cs6630 | October 23 2014

TEXT & SETS

Miriah Meyer
University of Utah

administrivia . . .

- parallel coordinates due next Thursday last time . . .

dataset types

Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

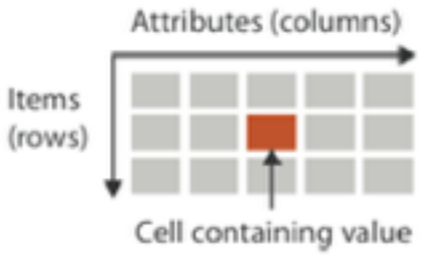
Geometry

Items

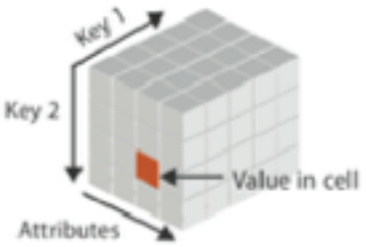
Positions

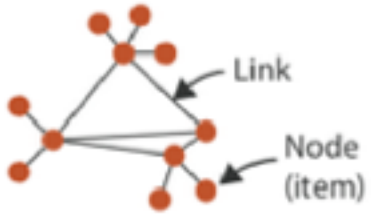
Clusters, Sets, Lists

Items

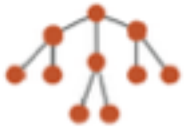


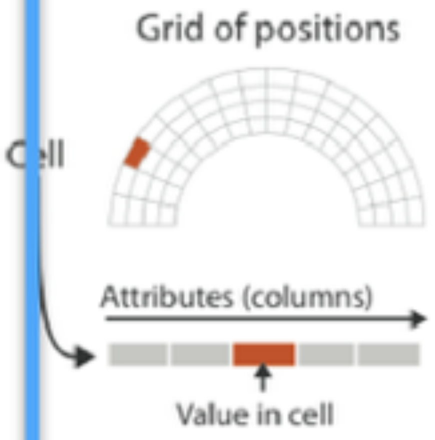
→ *Multidimensional Table*





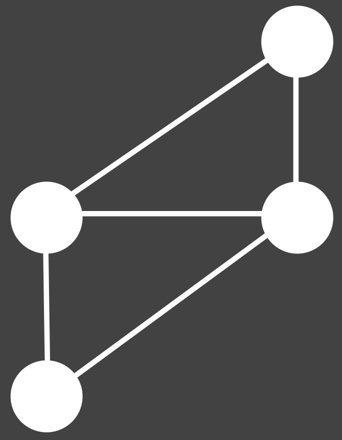
→ *Trees*



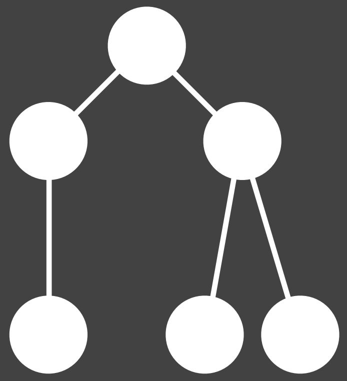


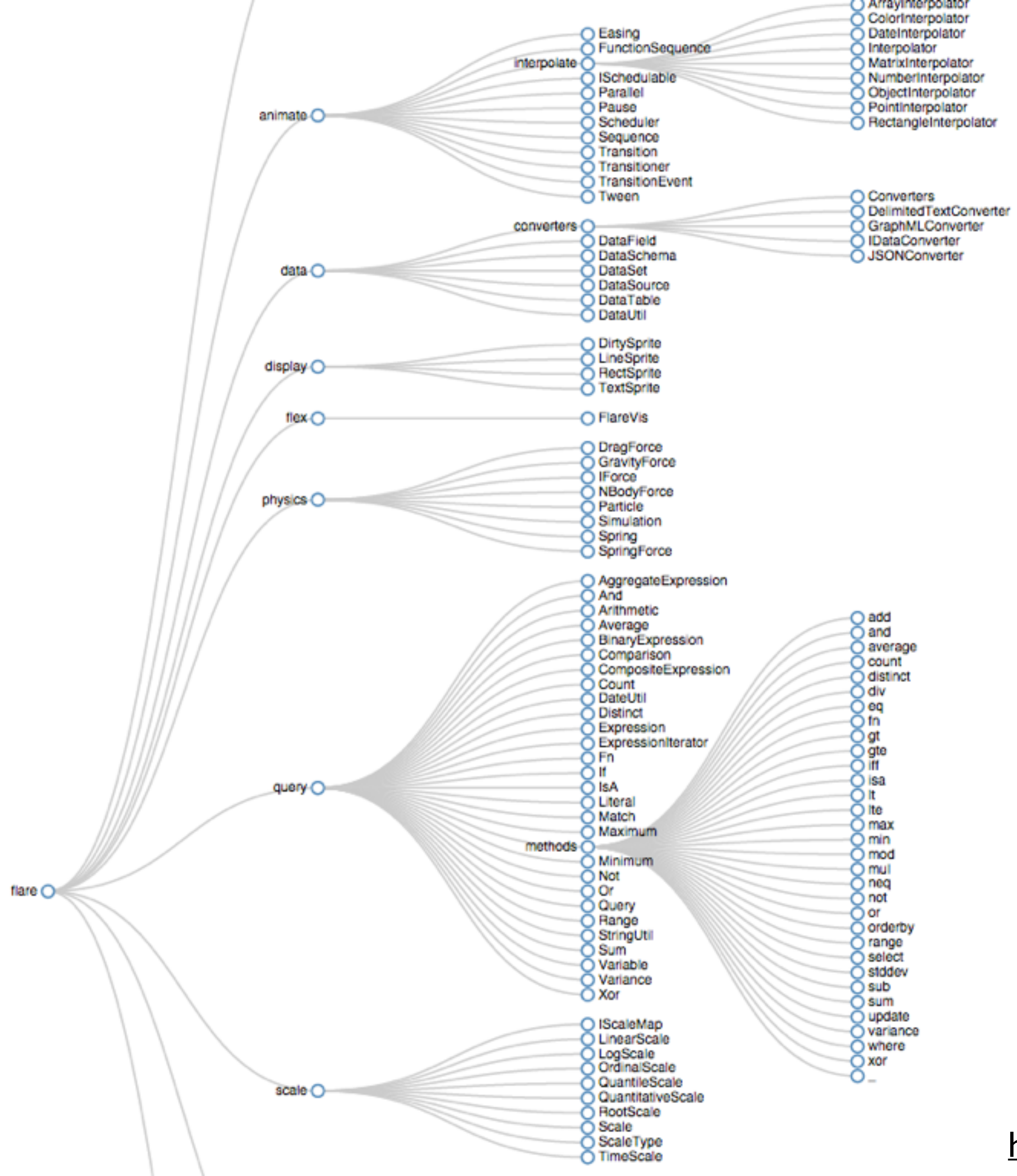
GRAPHS & TREES

- **graphs**
  - model relations amount data
  - *nodes* and *edges*

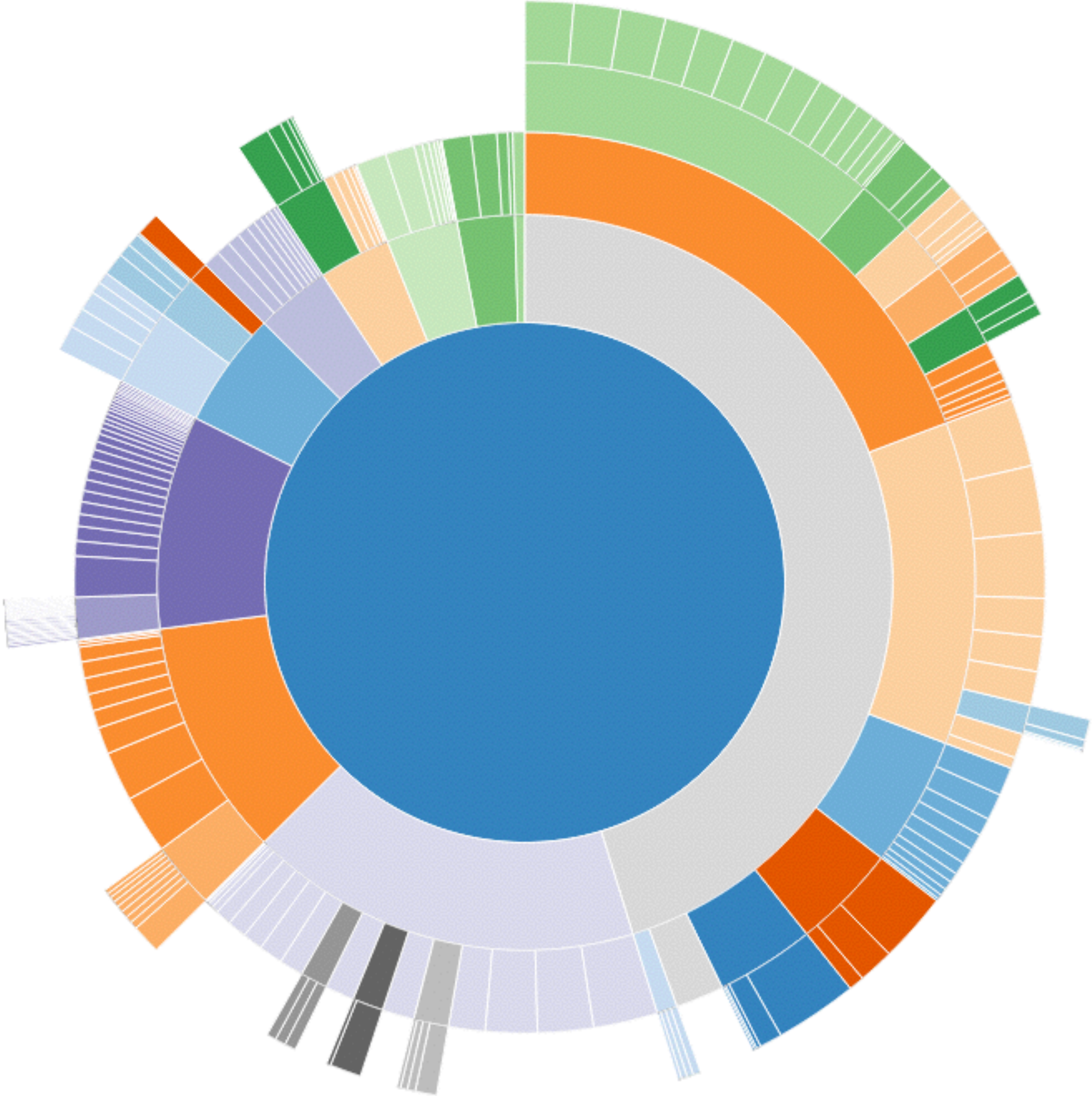


- **trees**
  - graphs with hierarchical structure
  - nodes as *parents* and *children*

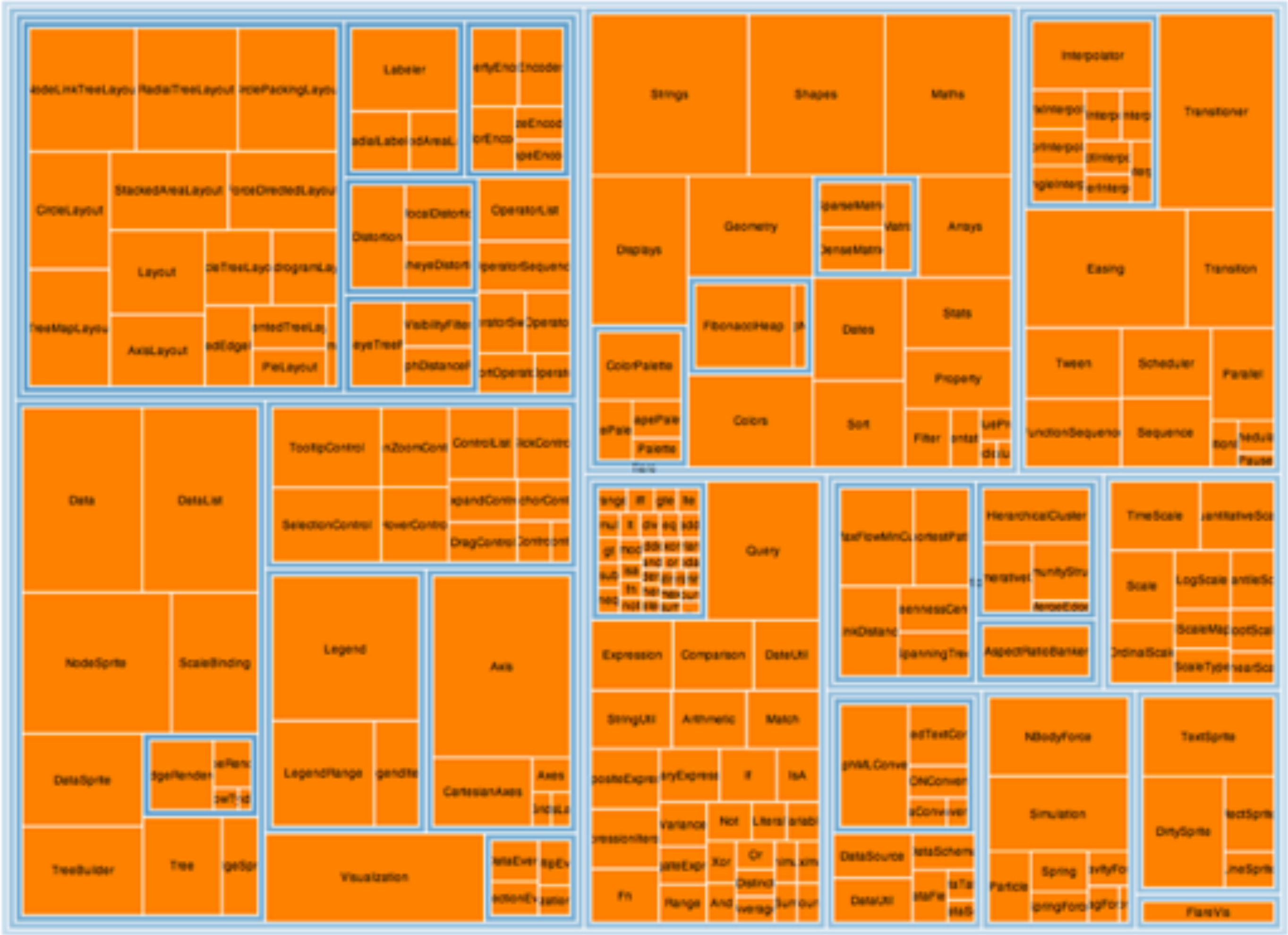


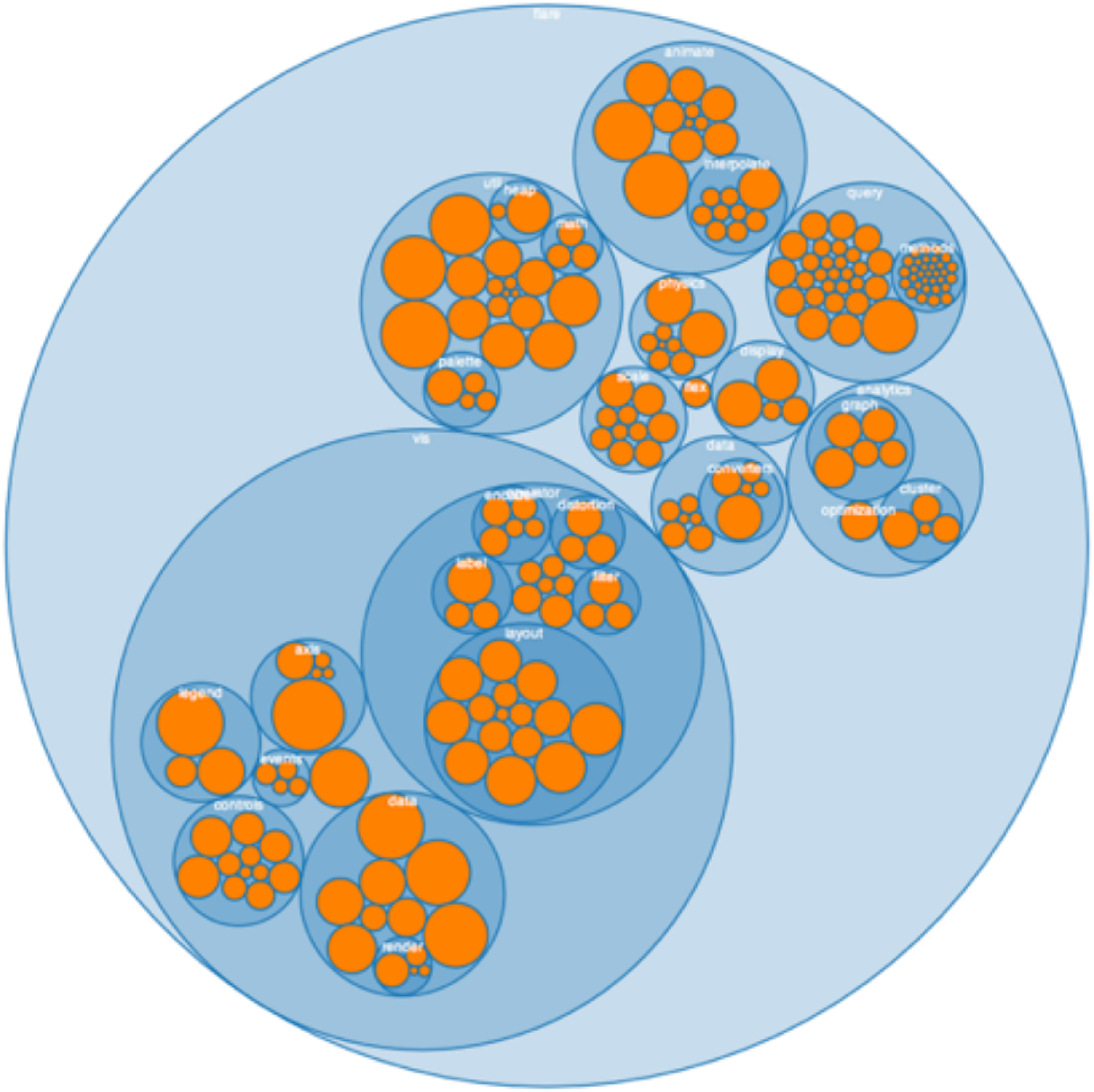


http://bl.ocks.org/mbostock/4339184



http://bl.ocks.org/mbostock/raw/4348373/





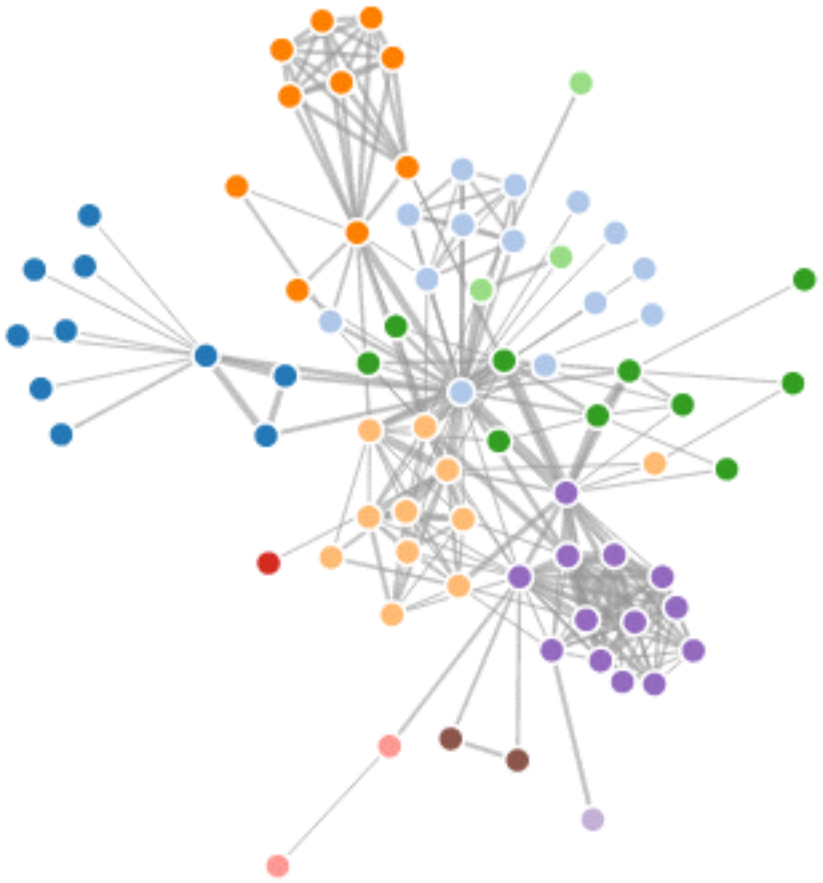
http://hci.stanford.edu/jheer/files/zoo/

VISUALIZING GRAPHS

- **node link layouts**
  - Reingold-Tilford (trees only)
  - Sugiyama (directed acyclic graphs)
  - Force directed
  - Attribute-based
- **adjacency matrices**
- **aggregate views**
  - Motif Glyphs
  - PivotGraph

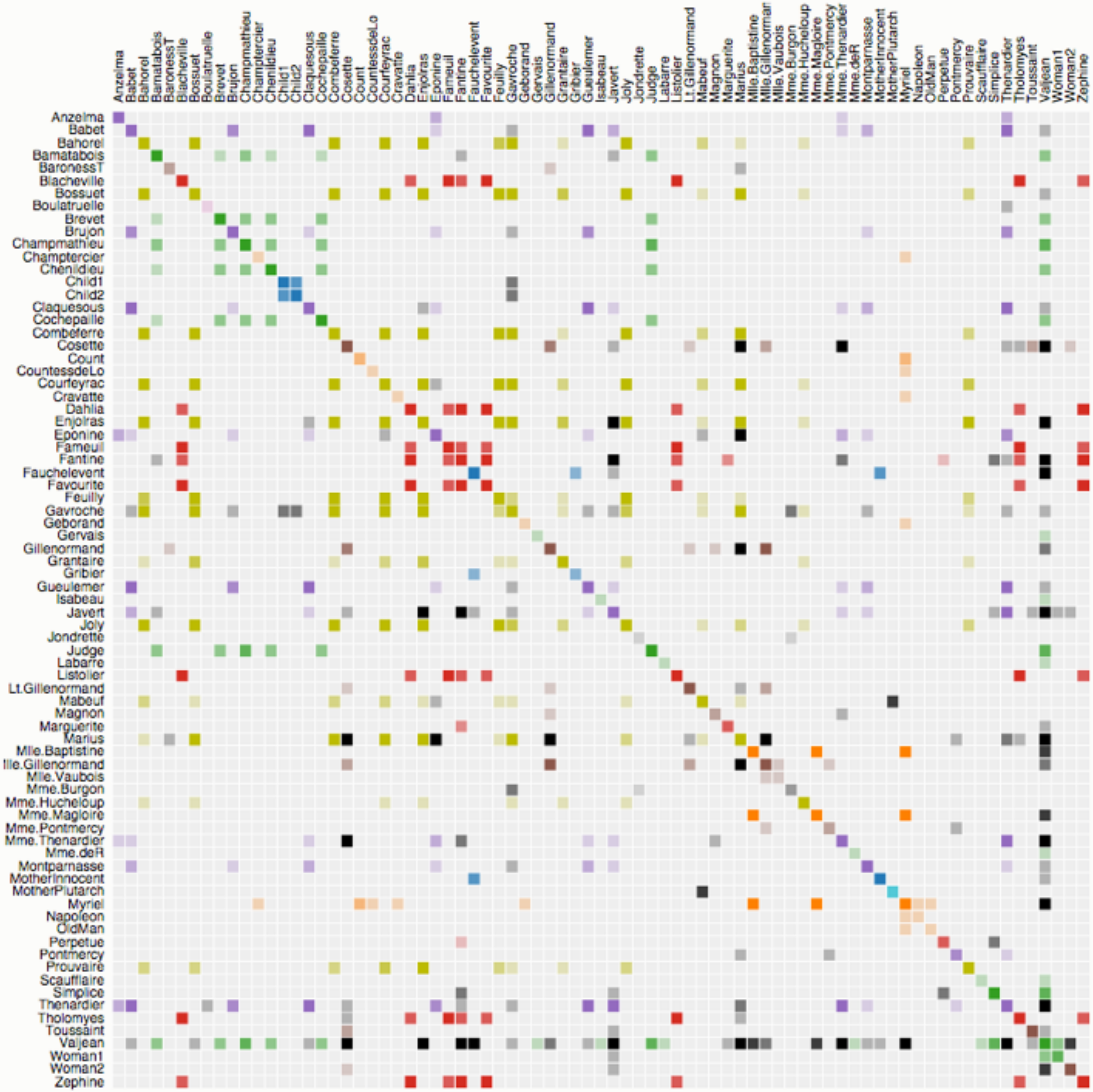
Les Misérables

character co-occurrence



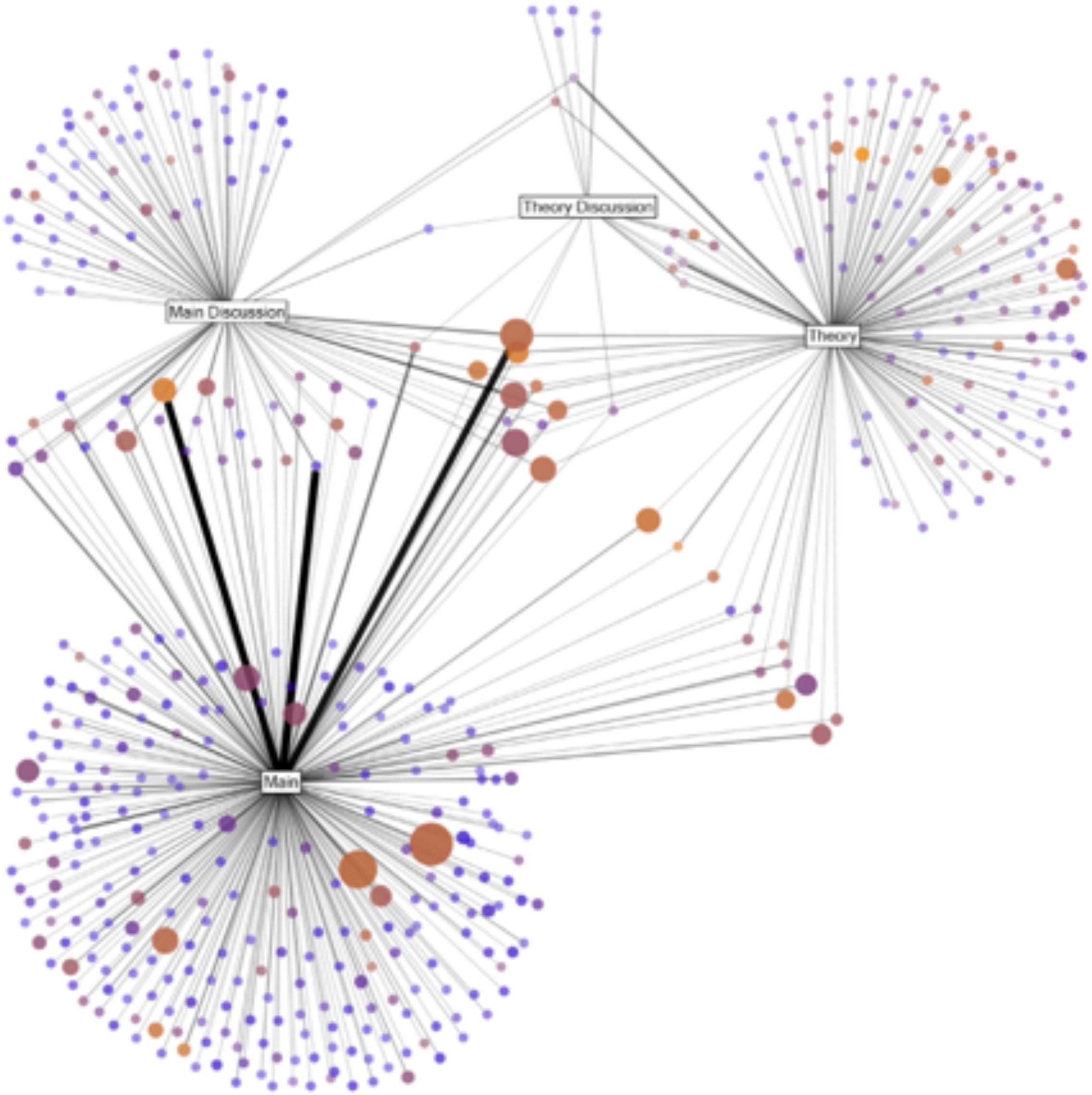
http://bl.ocks.org/mbostock/4062045

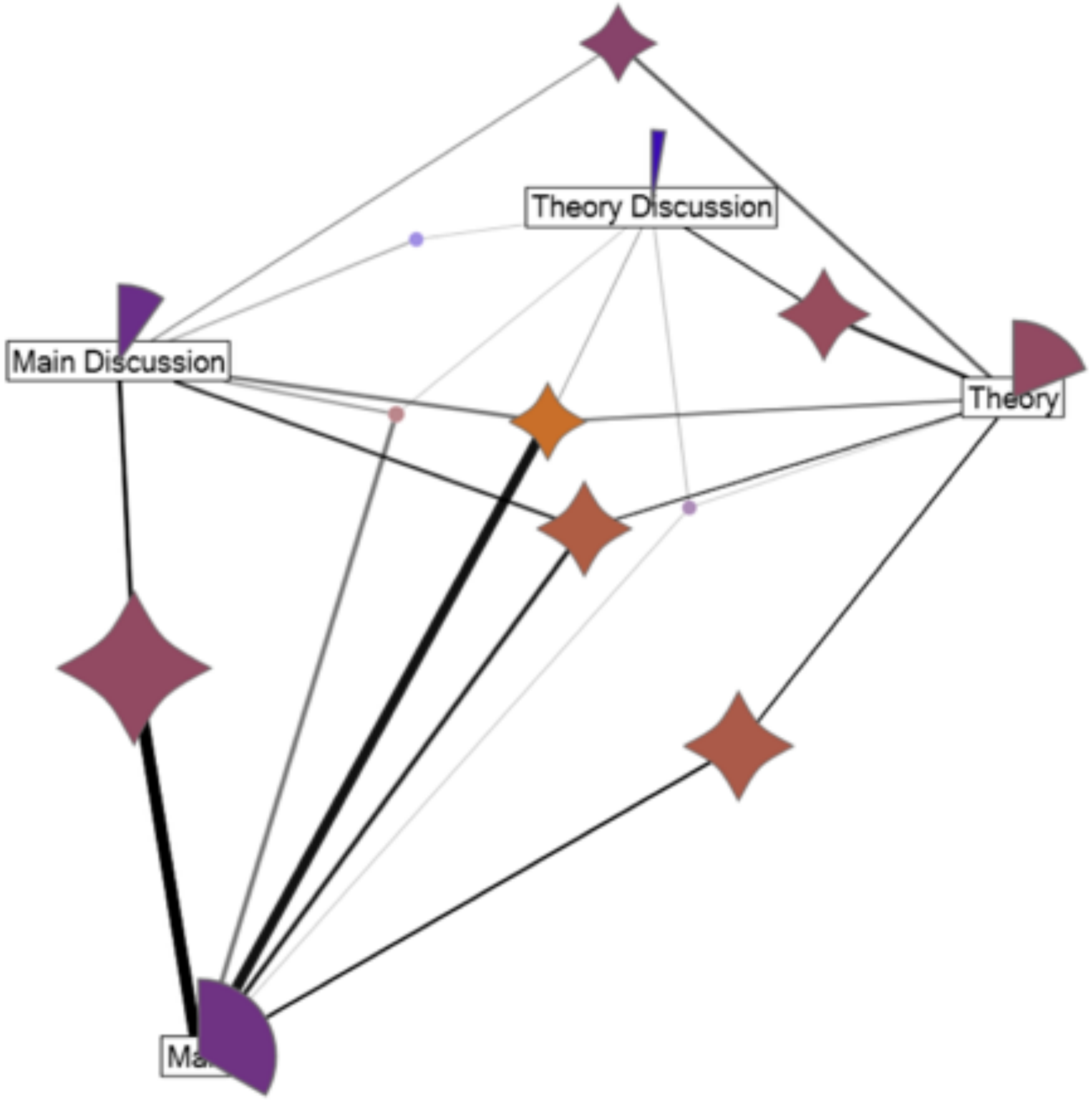
Les Misérables
character
co-occurrence



http://bost.ocks.org/mike/miserables/

MOTIF GLYPHS

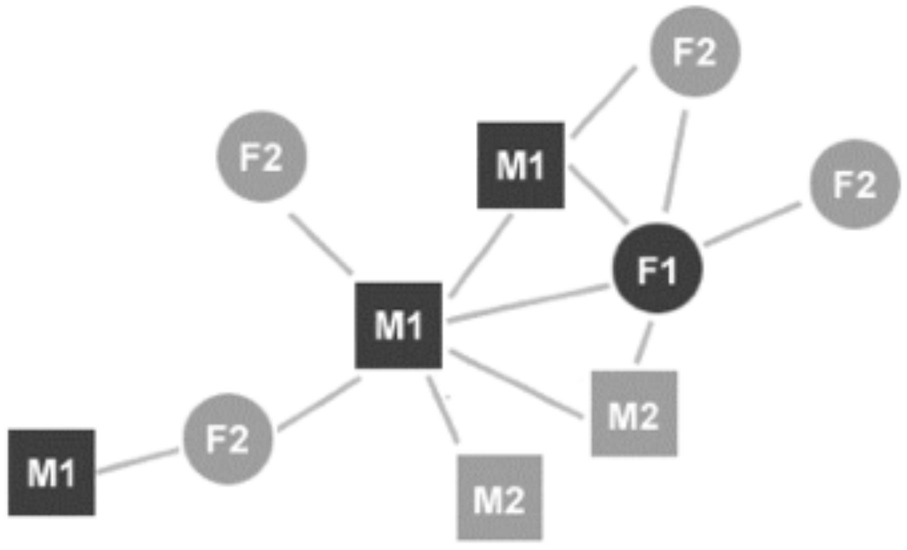




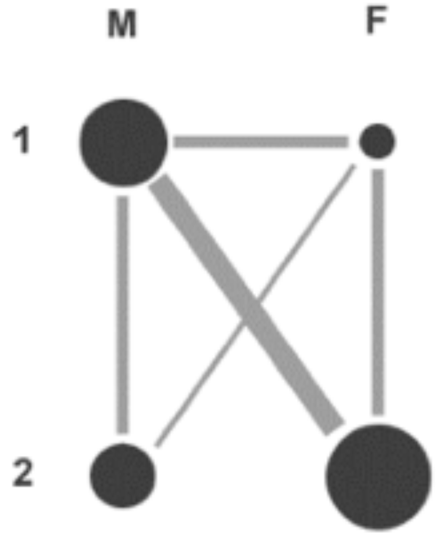
Dunne 2013

PIVOT GRAPHS

- new graph, derived from categorical node attributes
- 1D or 2D layouts possible
- size of nodes and edges related to number of aggregated original nodes and edges
- scalability through abstraction, not layout algorithm

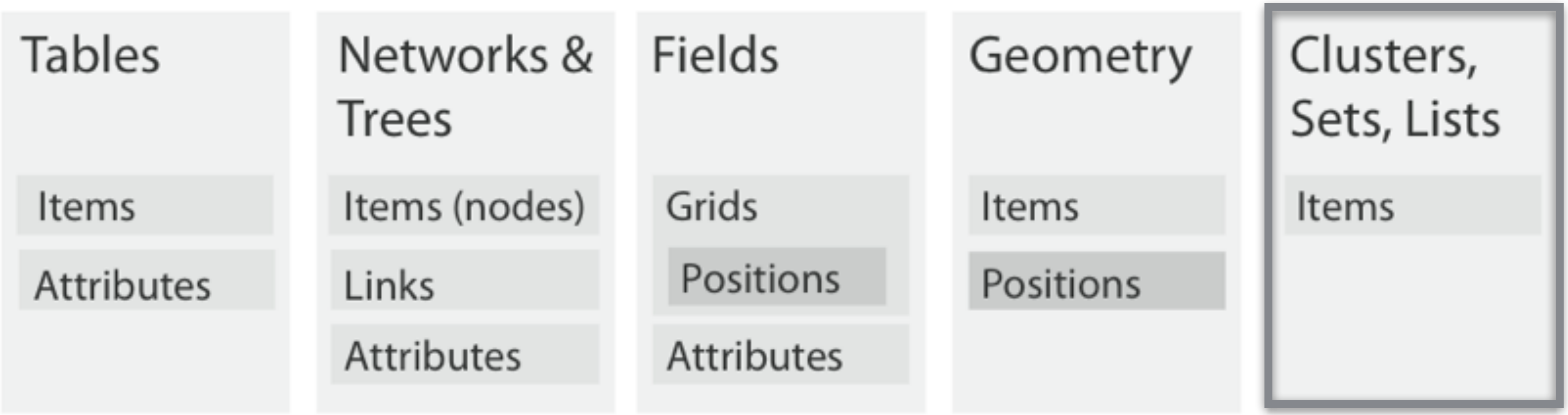


Node and Link Diagram

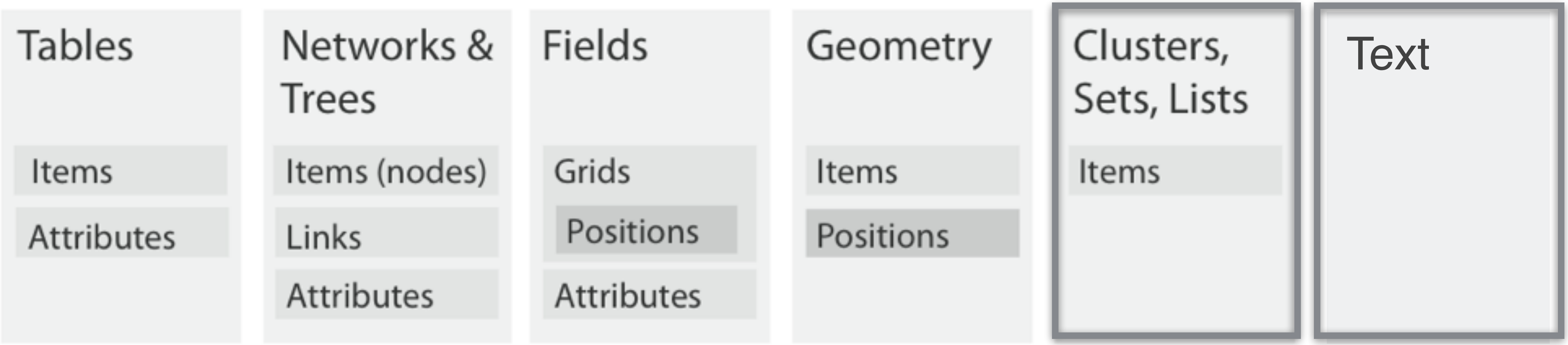


PivotGraph Roll-up

today . . .



Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		



Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

Geometry

Items

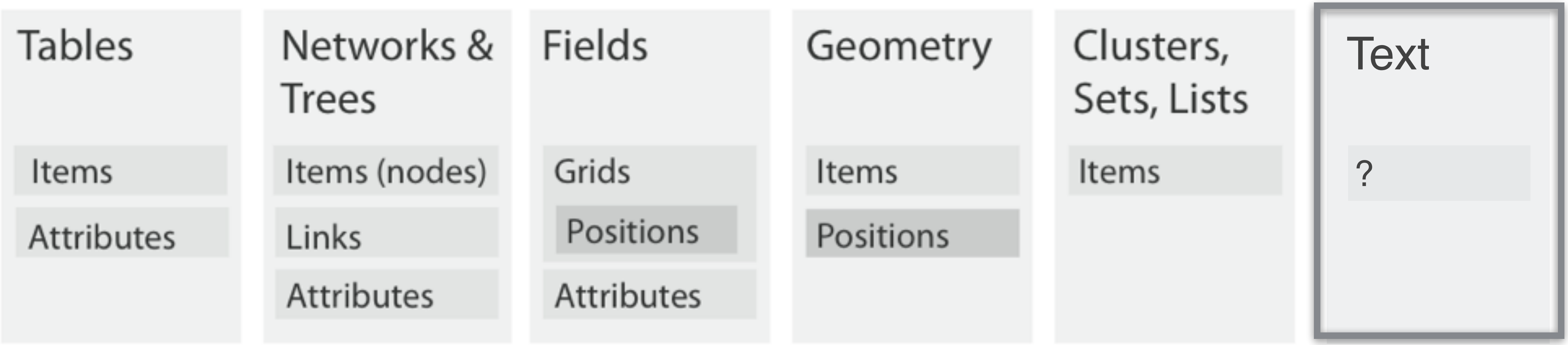
Positions

Clusters, Sets, Lists

Items

Text

TEXT



Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

Geometry

Items

Positions

Clusters, Sets, Lists

Items

Text

?

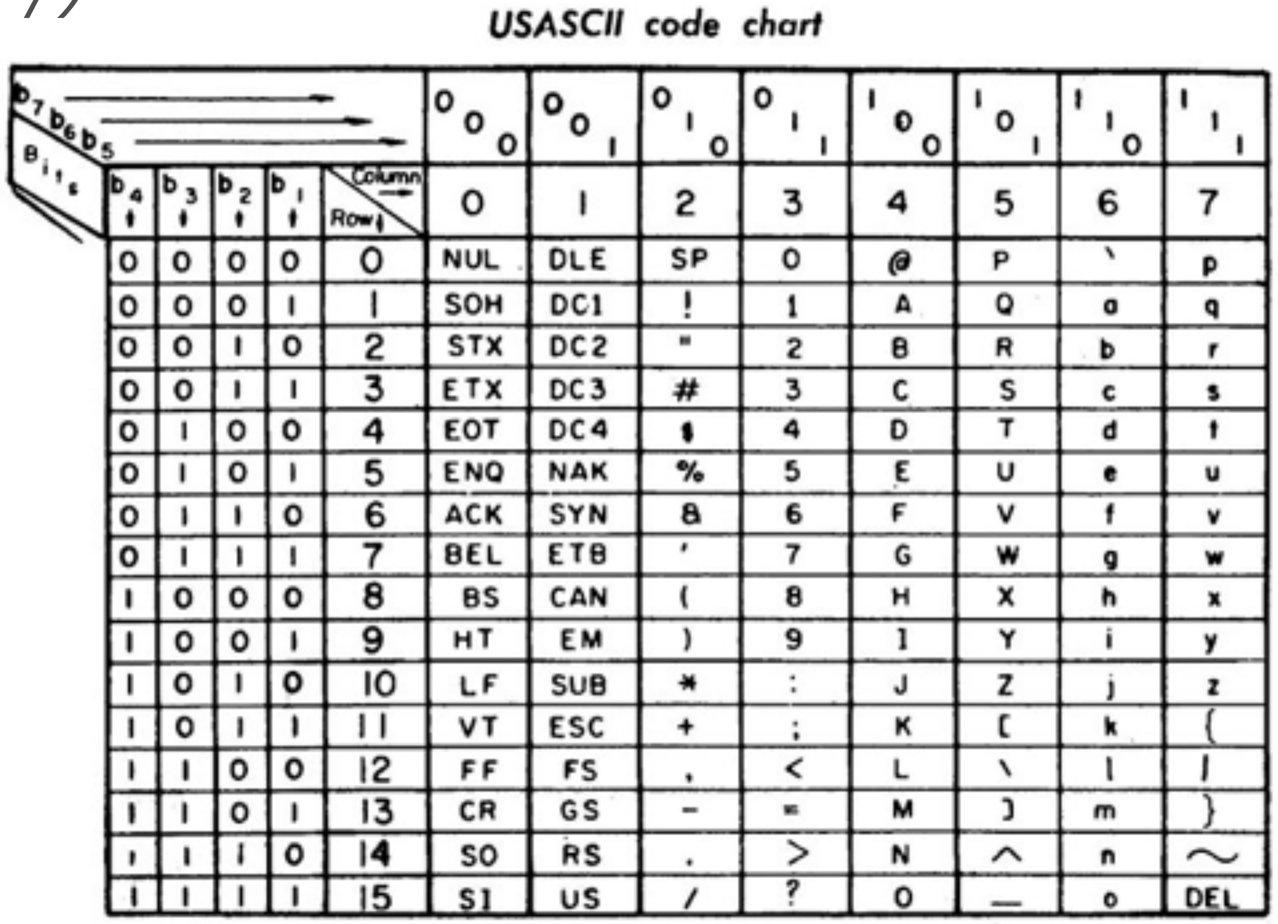
WHAT DOES IT MEAN TO BE AN "ITEM"?

text data type

- no numbers *(implicitly)*

- characters: ASCII

- strings



USASCII code chart

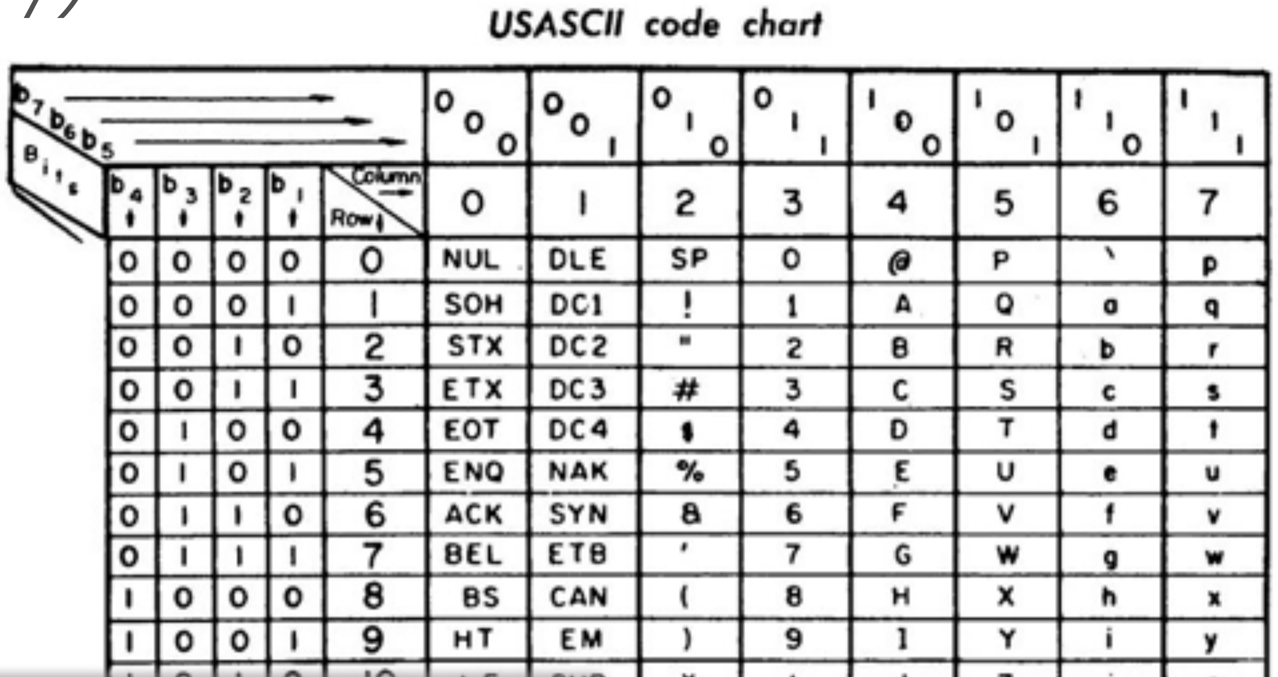
b7 b6 b5 →					0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b4	b3	b2	b1	Column → / Row ↓	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(	8	H	X	h	x
1	0	0	1	9	HT	EM	)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[	k	{
1	1	0	0	12	FF	FS	,	<	L	\	l	|
1	1	0	1	13	CR	GS	−	=	M	]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

text data type

- no numbers *(implicitly)*

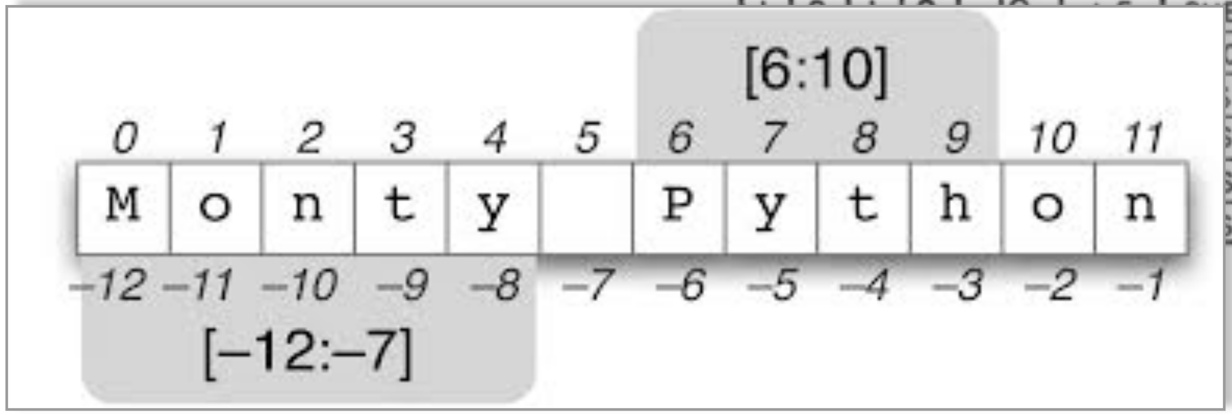
- characters: ASCII

- strings

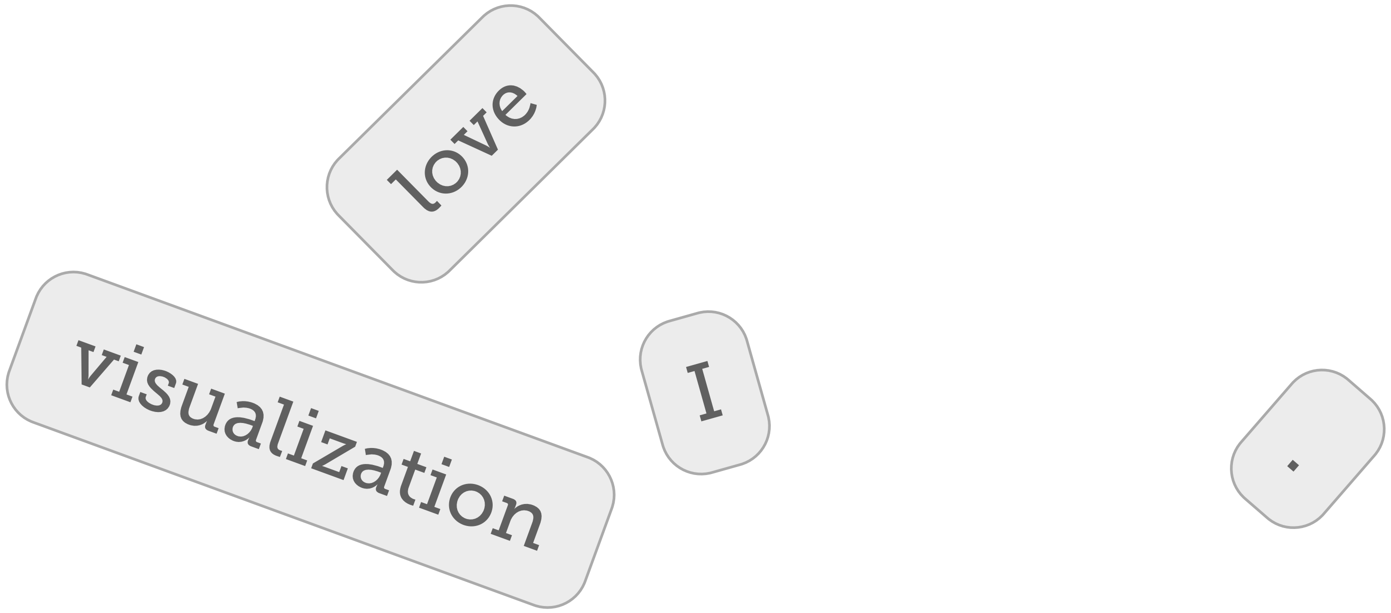


USASCII code chart

b7 b6 b5 →					0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
b4	b3	b2	b1	Column → / Row ↓	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(	8	H	X	h	x
1	0	0	1	9	HT	EM	)	9	I	Y	i	y
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	*	:	J	Z	j	z
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	+	;	K	[	k	{
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	,	<	L	\	l	|
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-	=	M	]	m	}
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.	>	N	^	n	~
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	/	?	O	_	o	DEL



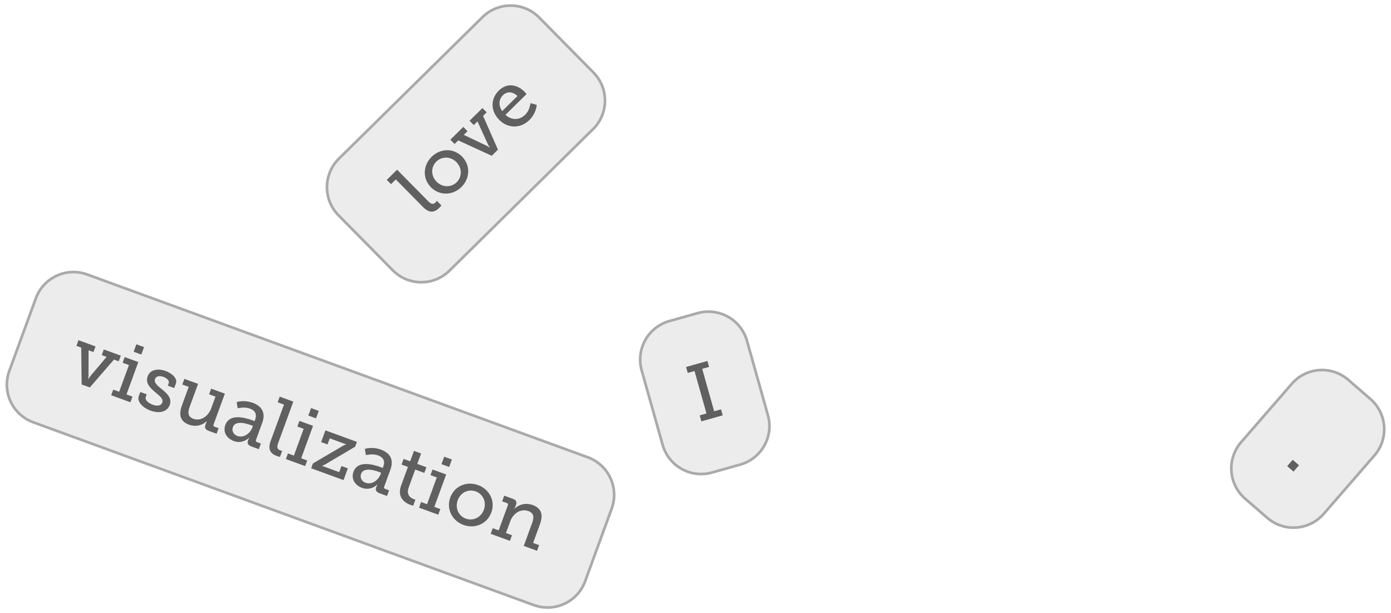
text data semantics



- words
- sentences
  - paragraphs
  - chapters
- lines

text data semantics

- words



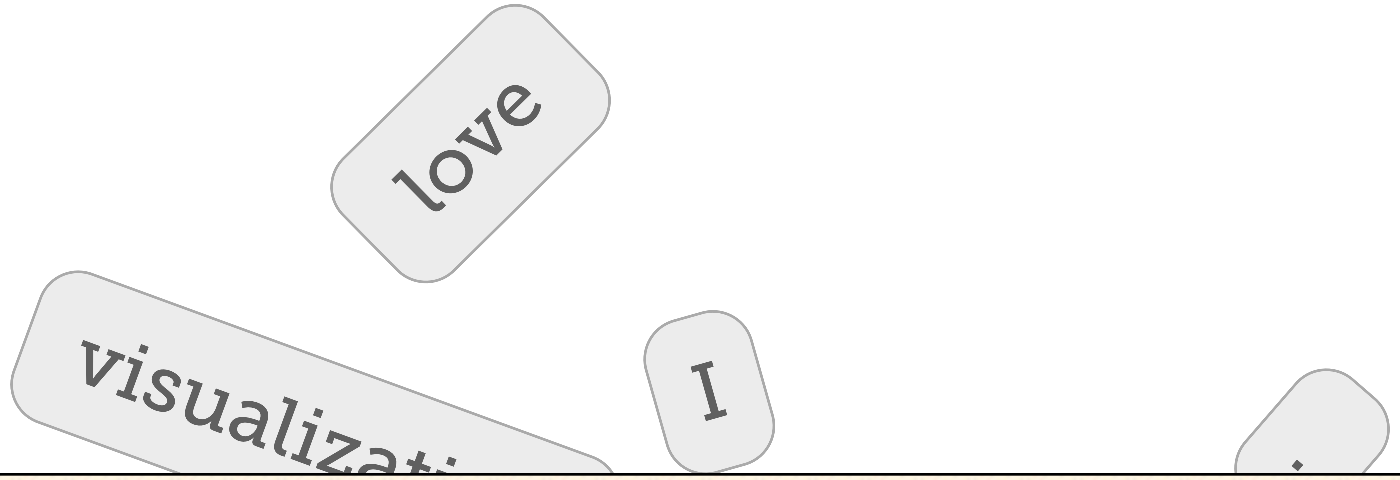
- sentences
  - paragraphs
  - chapters

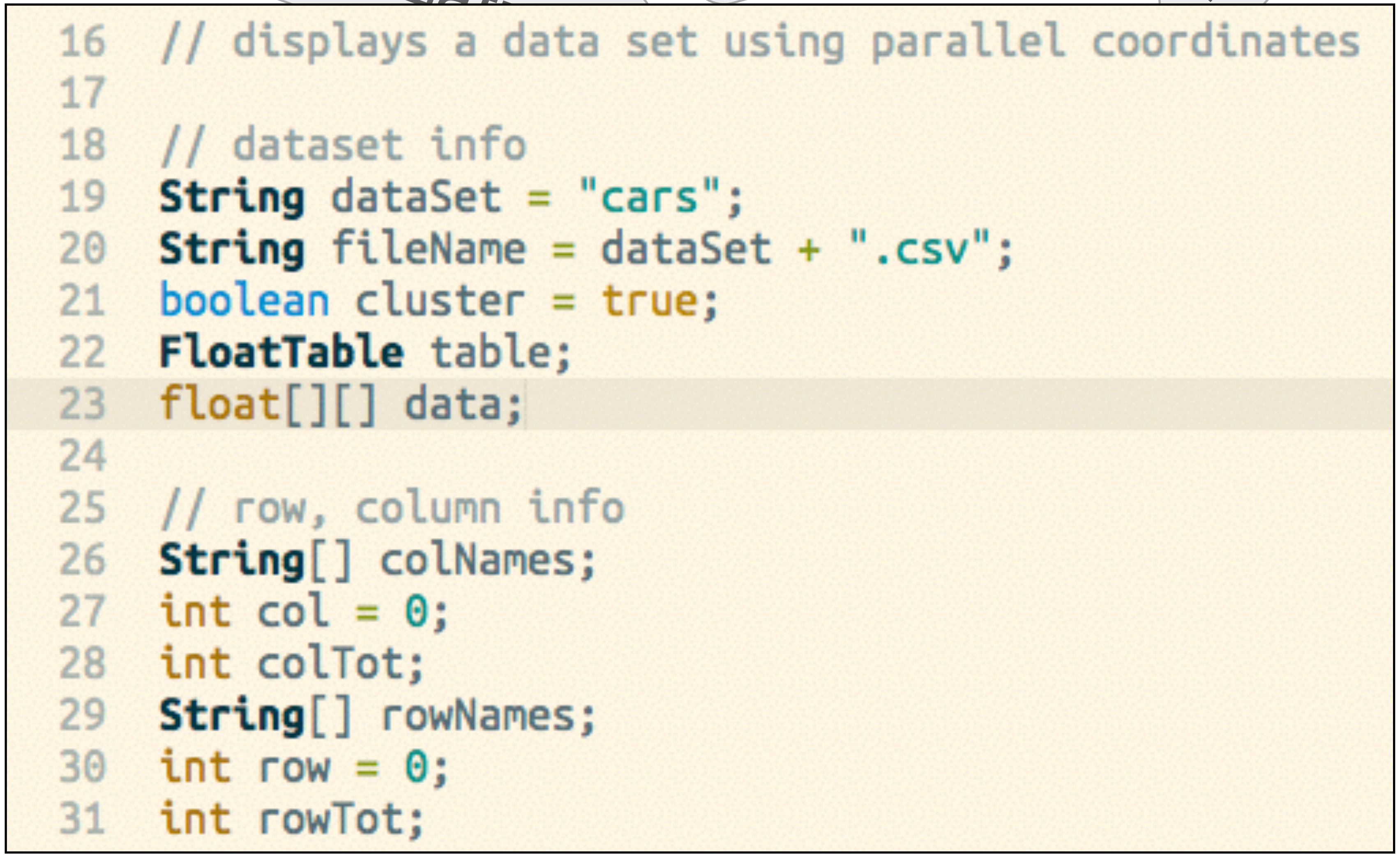
I love visualization.

- lines

text data semantics

- words
- sentences
  - paragraphs
  - chapters
- lines





```
// displays a data set using parallel coordinates

// dataset info
String dataSet = "cars";
String fileName = dataSet + ".csv";
boolean cluster = true;
FloatTable table;
float[][] data;

// row, column info
String[] colNames;
int col = 0;
int colTot;
String[] rowNames;
int row = 0;
int rowTot;
```

text data semantics

- **documents**
  - books
  - papers
  - webpages
  - emails
  - twitter post

Chapter 3
Automatic Mapping of Social Networks of Actors from Text Corpora: Time Series Analysis

Semantics-Based Analysis and Navigation of Heterogeneous Text Corpora: The *Porpoise* News and Blogs Engine

Automatic Analysis of Large Text Corpora - A Contribution to Structuring WEB Communities

Gerhard Heyer, Uwe Quasthoff, and Christian Wolff

Examining the Use of Text Corpora and Online Dictionaries as Learning Tools: Pre-Service Teachers' Perspectives

Constructing a Large Scale Text Corpus Based on the Grid and Trustworthiness

Searching Text Corpora with grep

AIML Knowledge Base Construction from Text Corpora

- **corpus:** *collection of documents*

text data semantics

- **documents**
  - books
  - papers
  - webpages
  - emails
  - twitter post

Chapter 3
Automatic Mapping of Social Networks of Actors from Text Corpora: Time Series Analysis

Semantics-Based Analysis and Navigation of Heterogeneous Text Corpora: The *Porpoise* News and Blogs Engine

Automatic Analysis of Large Text Corpora - A Contribution to Structuring WEB Communities

Examining the Use of
and Online Dictionari
Pre-Service Teachers'

Chelsea Young @YoungChelsMarie 3h
@EliYoungBand can't wait to see you guys here in Buffalo,NY at the @1065wyrk Taste of Country in June! PS-My daddy's name is James Young! :)

Kat Przybyla @katprz 5h
Who's coming to the @1065wyrk Taste of Country? A trifecta of southern hotties -- @ericchurch @JoeNichols @EliYoungBand buffalo.com/entertainment/...

Alica Wiedenbeck @leesh06 6h
@Eric_Church at @1065wyrk Taste of Country this year..... dying. #Cantwait

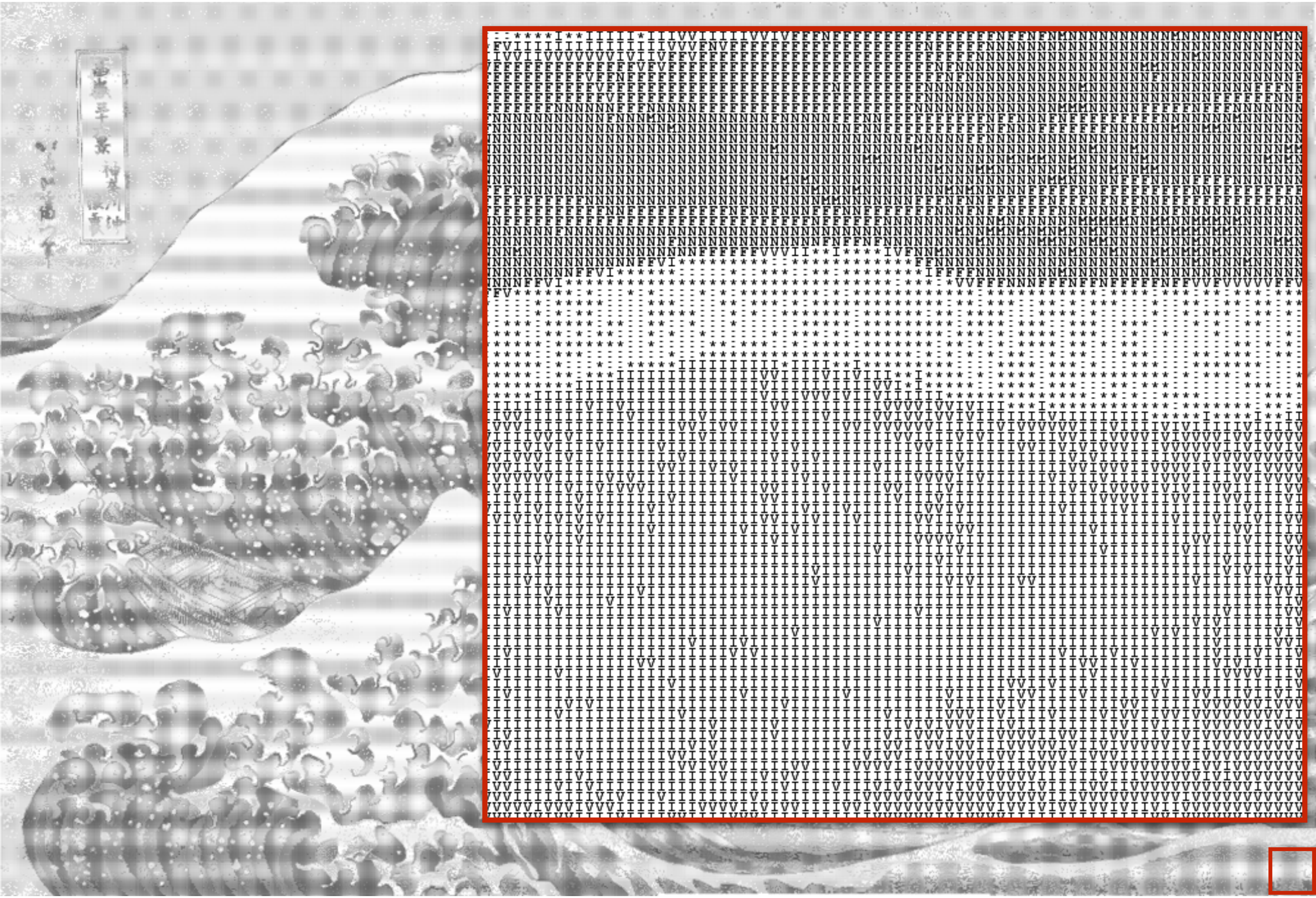
Michele L. McDaniel @michelemcd13 8h
absolutely thrilled that @ericchurch will be headlining the @1065wyrk taste of country! Only 135 days! (but who's counting?!)

Jaime Lynn @JaimeLynn478 9h
@ClayModen @1065wyrk SOOOOOOO excited for TOC this year...def bringing my 5yr old, we are HUGE Eric Church fans!!!

Jake Rosen @JRosen26 10h
To say @1065wyrk got taste right this year would be an understatement #ericchurch

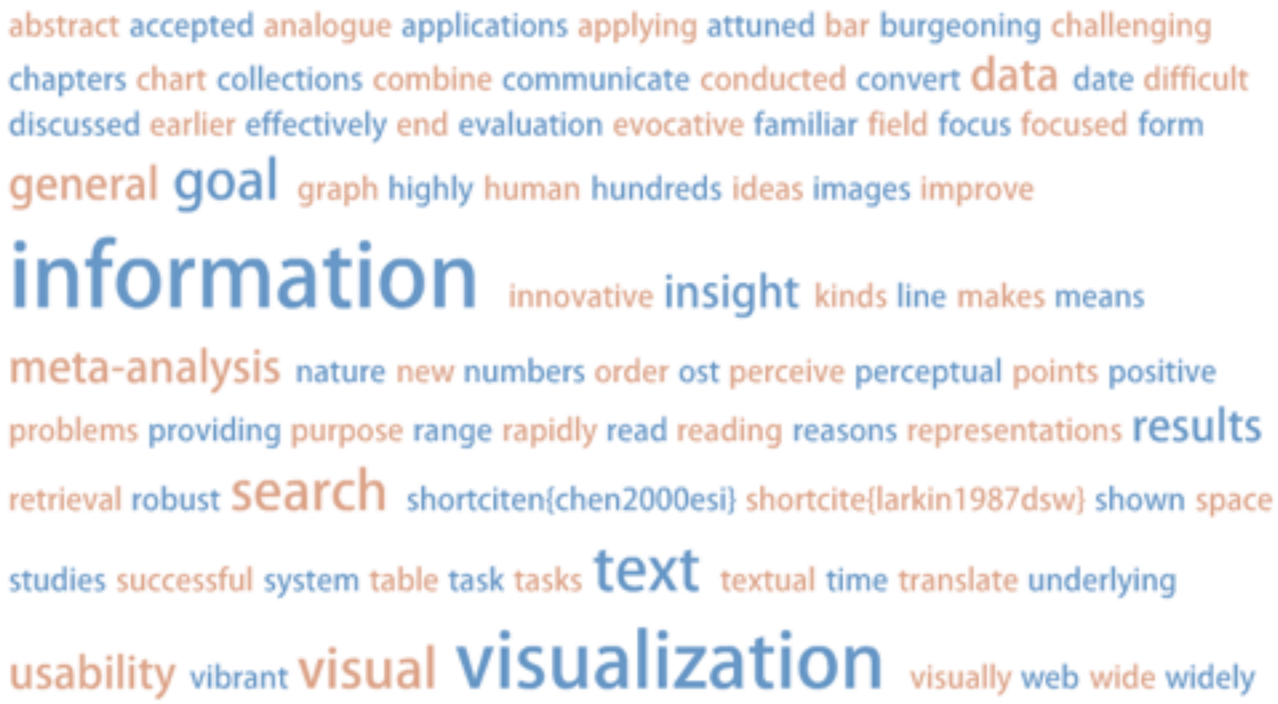
- **corpus:** *collection of documents*

冨嶽三十六景 神奈川沖 浪裏
北斎改為一筆

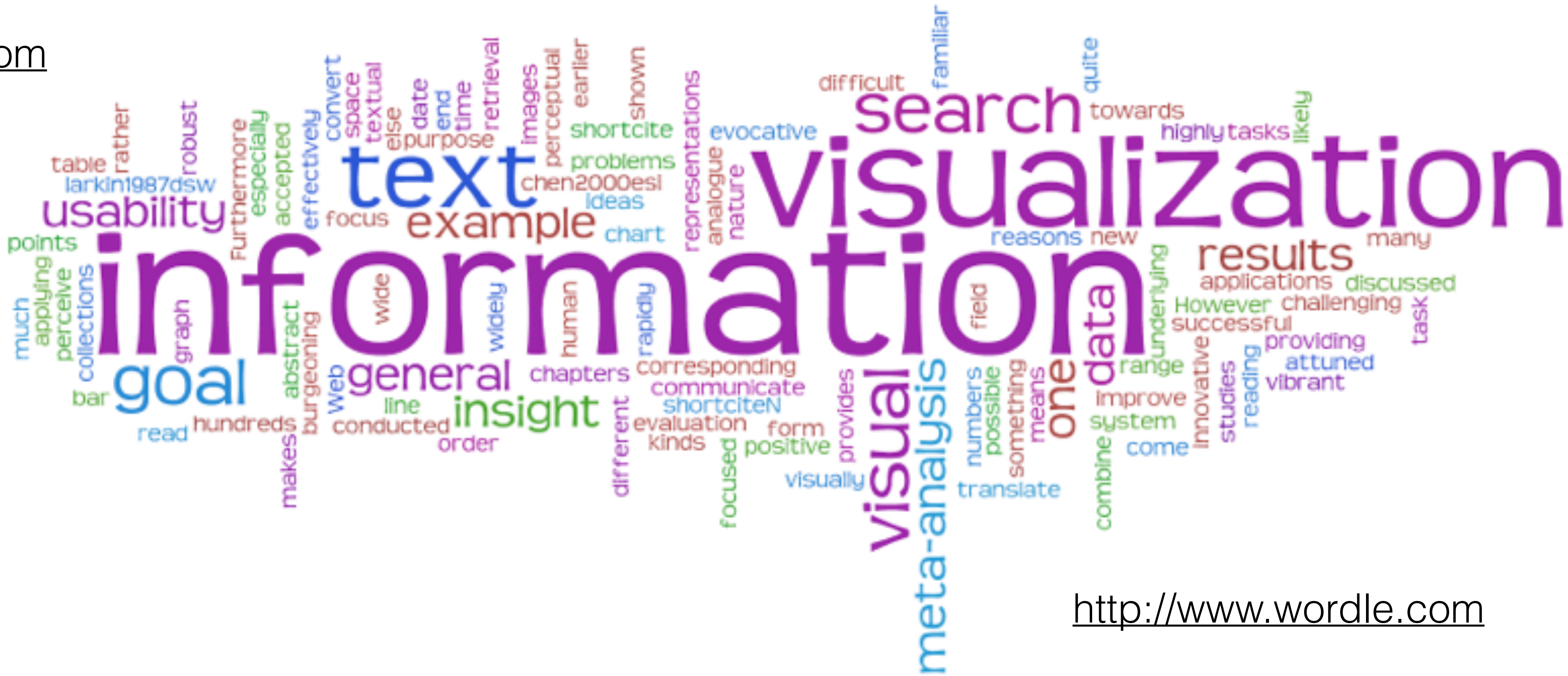


single document

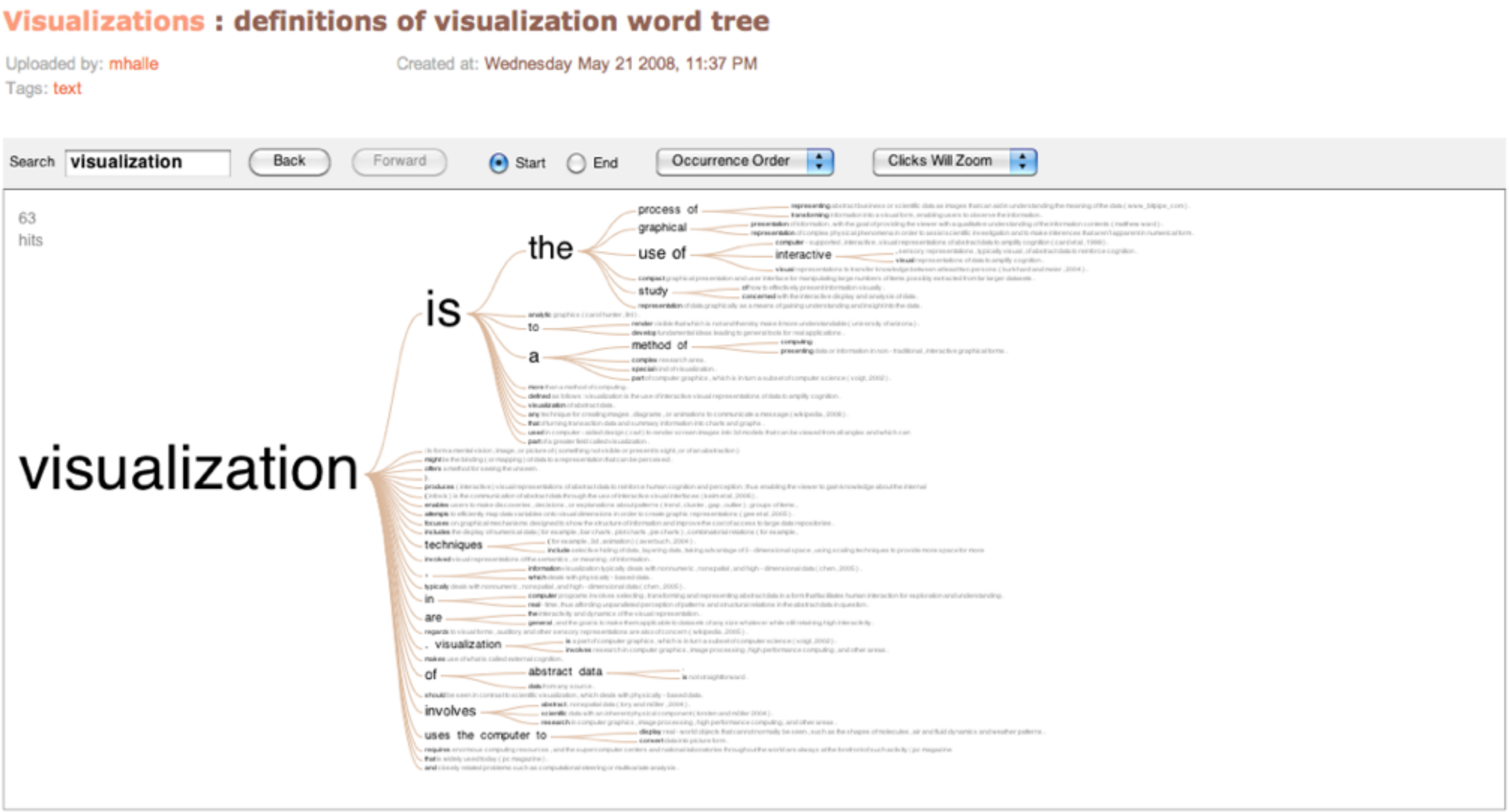
Tag Clouds / Word Clouds


abstract accepted analogue applications applying attuned bar burgeoning challenging
chapters chart collections combine communicate conducted convert data date difficult
discussed earlier effectively end evaluation evocative familiar field focus focused form
general goal graph highly human hundreds ideas images improve
information innovative insight kinds line makes means
meta-analysis nature new numbers order ost perceive perceptual points positive
problems providing purpose range rapidly read reading reasons representations results
retrieval robust search shortciten{chen2000esi} shortcite{larkin1987dsw} shown space
studies successful system table task tasks text textual time translate underlying
usability vibrant visual visualization visually web wide widely

http://www.tagcrowd.com



http://www.wordle.com



Visualizations : definitions of visualization word tree

Uploaded by: mhalle

Created at: Wednesday May 21 2008, 11:37 PM

Tags: text

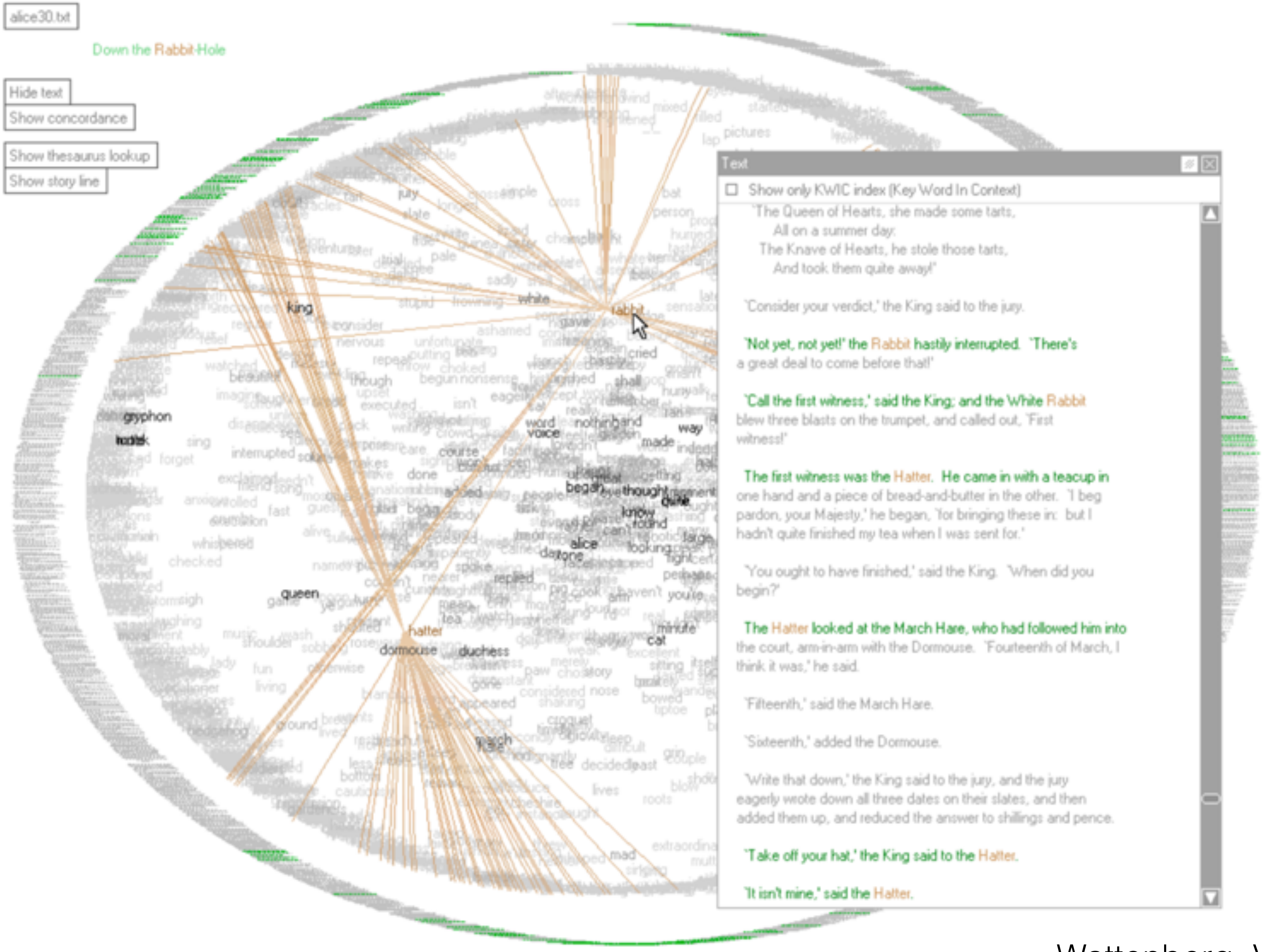
Search: visualization

Back | Forward | ● Start ○ End | Occurrence Order | Clicks Will Zoom

63 hits

visualization
is
the
process of
representing abstract business or scientific data as images that can aid in understanding the meaning of the data (www_bitpipe_com) .
transforming information into a visual form, enabling users to observe the information .
graphical
presentation of information, with the goal of providing the viewer with a qualitative understanding of the information contents (mathew ward) .
representation of complex physical phenomena in order to assist scientific investigation and to make inferences that aren't apparent in numerical form .
use of
computer - supported , interactive , visual representations of abstract data to amplify cognition (card et al , 1999) .
interactive
, sensory representations , typically visual , of abstract data to reinforce cognition .
visual representations of data to amplify cognition .
visual representations to transfer knowledge between at least two persons (burkhard and meier , 2004) .
compact graphical presentation and user interface for manipulating large numbers of items possibly extracted from far larger datasets .
study
of how to effectively present information visually .
concerned with the interactive display and analysis of data .
representation of data graphically as a means of gaining understanding and insight into the data .
analytic graphics (carol hunter , llnl) .
to
render visible that which is not and thereby make it more understandable (university of arizona) .
develop fundamental ideas leading to general tools for real applications .
a
method of
computing .
presenting data or information in non - traditional , interactive graphical forms .
complex research area .
special kind of visualization .
part of computer graphics , which is in turn a subset of computer science (voigt , 2002) .
more than a method of computing .
defined as follows : visualization is the use of interactive visual representations of data to amplify cognition .
visualization of abstract data .
any technique for creating images , diagrams , or animations to communicate a message (wikipedia , 2006) .
that of turning transaction data and summary information into charts and graphs .
used in computer - aided design (cad) to render screen images into 3d models that can be viewed from all angles and which can
part of a greater field called visualization .
is to form a mental vision , image , or picture of (something not visible or present to sight , or of an abstraction)
might be the binding (or mapping) of data to a representation that can be perceived .
offers a method for seeing the unseen .
)
produces (interactive) visual representations of abstract data to reinforce human cognition and perception , thus enabling the viewer to gain knowledge about the internal
(infovis) is the communication of abstract data through the use of interactive visual interfaces (keim et al , 2006) .
enables users to make discoveries , decisions , or explanations about patterns (trend , cluster , gap , outlier) , groups of items .
attempts to efficiently map data variables onto visual dimensions in order to create graphic representations (gee et al , 2005) .
focuses on graphical mechanisms designed to show the structure of information and improve the cost of access to large data repositories .
includes the display of numerical data (for example , bar charts , plot charts , pie charts) , combinatorial relations (for example ,
techniques
(for example , 3d , animation) (averbuch , 2004) .
include selective hiding of data , layering data , taking advantage of 3 - dimensional space , using scaling techniques to provide more space for more
involved visual representations of the semantics , or meaning , of information .
,
information visualization typically deals with nonnumeric , nonspatial , and high - dimensional data (chen , 2005) .
which deals with physically - based data .
typically deals with nonnumeric , nonspatial , and high - dimensional data (chen , 2005) .
in
computer programs involves selecting , transforming and representing abstract data in a form that facilitates human interaction for exploration and understanding .
real - time , thus affording unparalleled perception of patterns and structural relations in the abstract data in question .
are
the interactivity and dynamics of the visual representation .
general , and the goal is to make them applicable to datasets of any size whatever while still retaining high interactivity .
regards to visual forms , auditory and other sensory representations are also of concern (wikipedia , 2005) .
. visualization
is a part of computer graphics , which is in turn a subset of computer science (voigt , 2002) .
involves research in computer graphics , image processing , high performance computing , and other areas .
makes use of what is called external cognition .
of
abstract data
.
is not straightforward .
data from any source .
should be seen in contrast to scientific visualization , which deals with physically - based data .
involves
abstract , nonspatial data (fory and miller , 2004) .
scientific data with an inherent physical component (tondon and miller 2004) .
research in computer graphics , image processing , high performance computing , and other areas .
uses the computer to
display real - world objects that cannot normally be seen , such as the shapes of molecules , air and fluid dynamics and weather patterns .
convert data into picture form .
requires enormous computing resources , and the supercomputer centers and national laboratories throughout the world are always at the forefront of such activity (pc magazine
that is widely used today (pc magazine) .
and closely related problems such as computational steering or multivariate analysis .

Text Arc



☐ Show only KWIC index (Key Word In Context)

'The Queen of Hearts, she made some tarts,
All on a summer day:
The Knave of Hearts, he stole those tarts,
And took them quite away!'

'Consider your verdict,' the King said to the jury.

'Not yet, not yet!' the Rabbit hastily interrupted. 'There's a great deal to come before that!'

'Call the first witness,' said the King; and the White Rabbit blew three blasts on the trumpet, and called out, 'First witness!'

The first witness was the Hatter. He came in with a teacup in one hand and a piece of bread-and-butter in the other. 'I beg pardon, your Majesty,' he began, 'for bringing these in: but I hadn't quite finished my tea when I was sent for.'

'You ought to have finished,' said the King. 'When did you begin?'

The Hatter looked at the March Hare, who had followed him into the court, arm-in-arm with the Dormouse. 'Fourteenth of March, I think it was,' he said.

'Fifteenth,' said the March Hare.

'Sixteenth,' added the Dormouse.

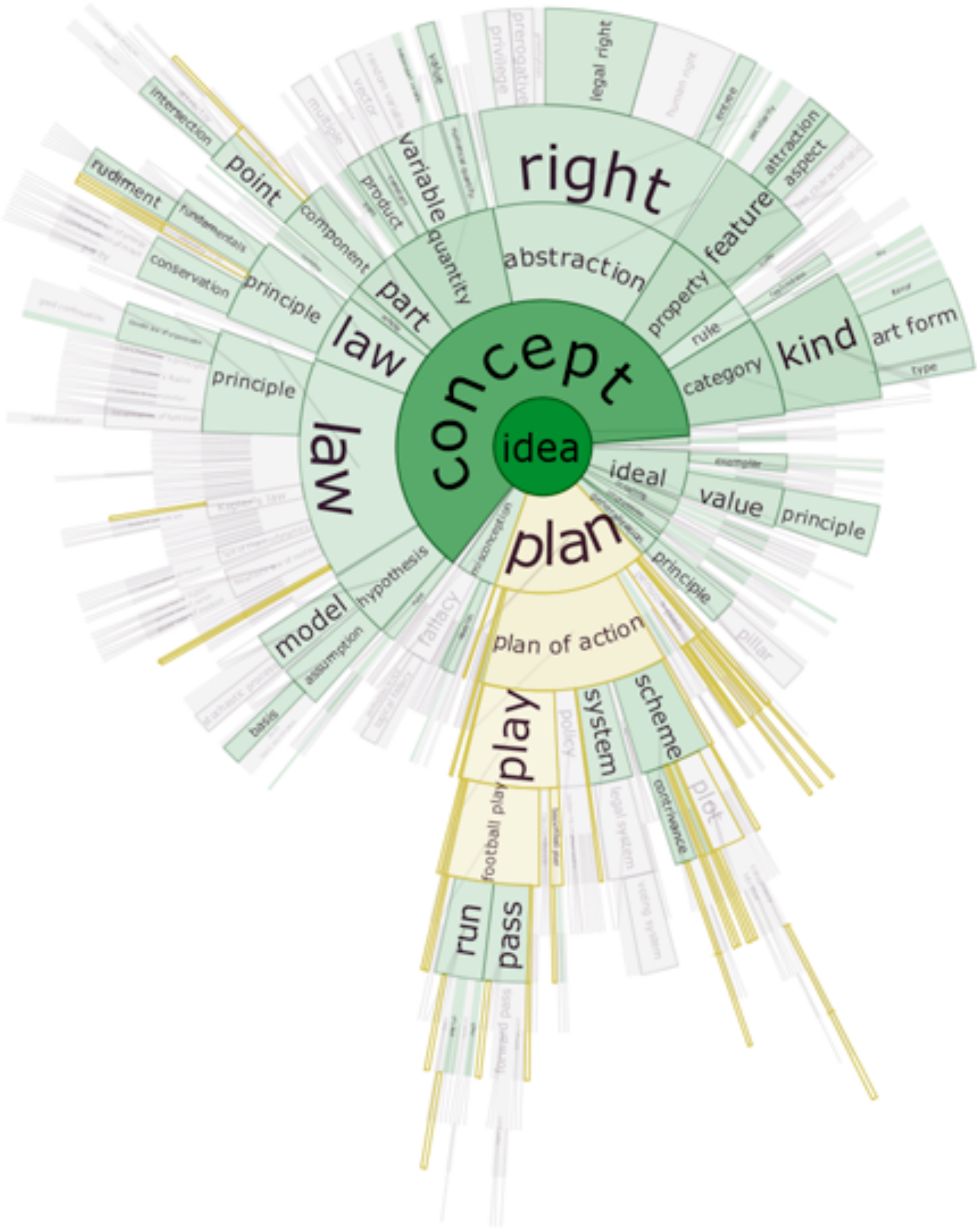
'Write that down,' the King said to the jury, and the jury eagerly wrote down all three dates on their slates, and then added them up, and reduced the answer to shillings and pence.

'Take off your hat,' the King said to the Hatter.

'It isn't mine,' said the Hatter.

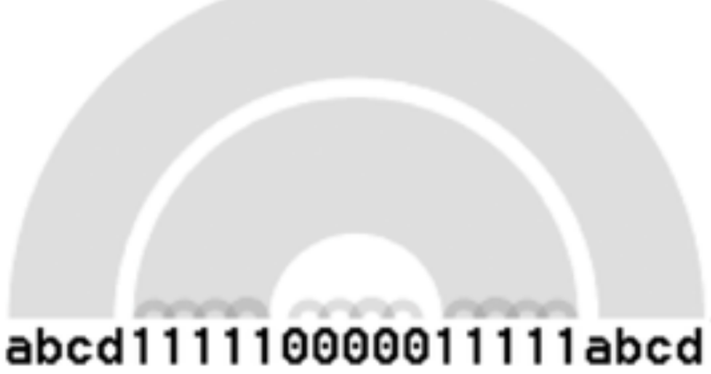
Wattenberg, Viegas 2008

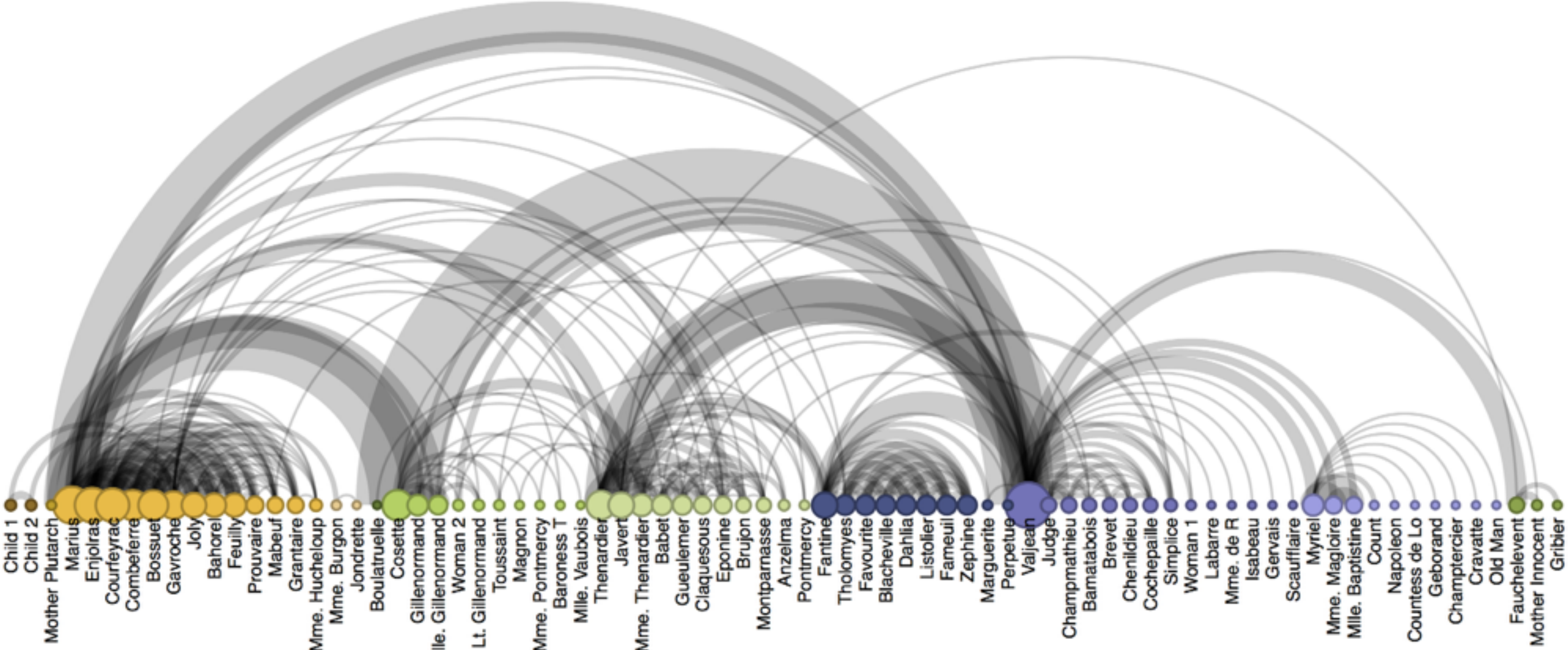
DocuBurst



Collins, Carpendale, Penn 2008

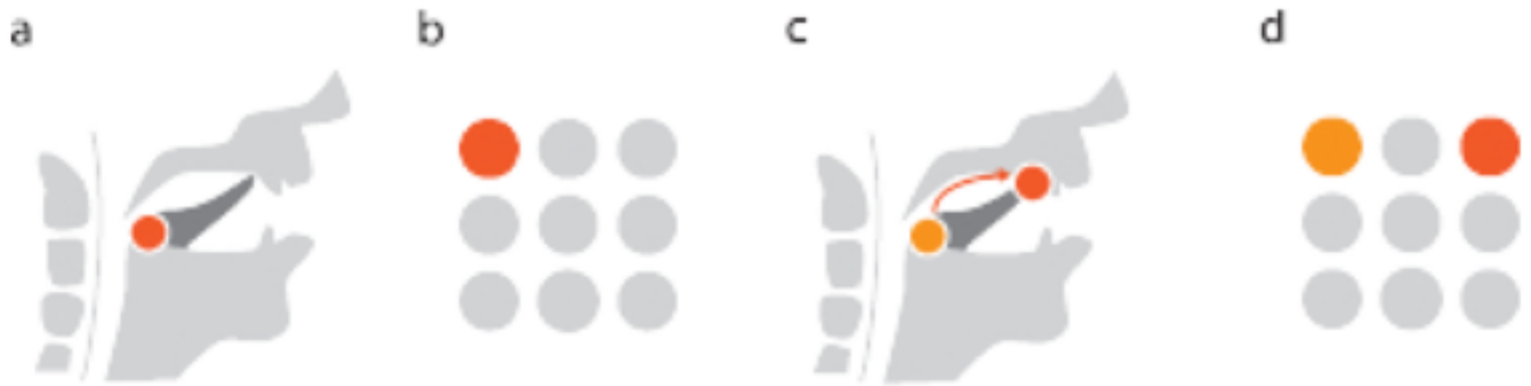
Arc Diagrams

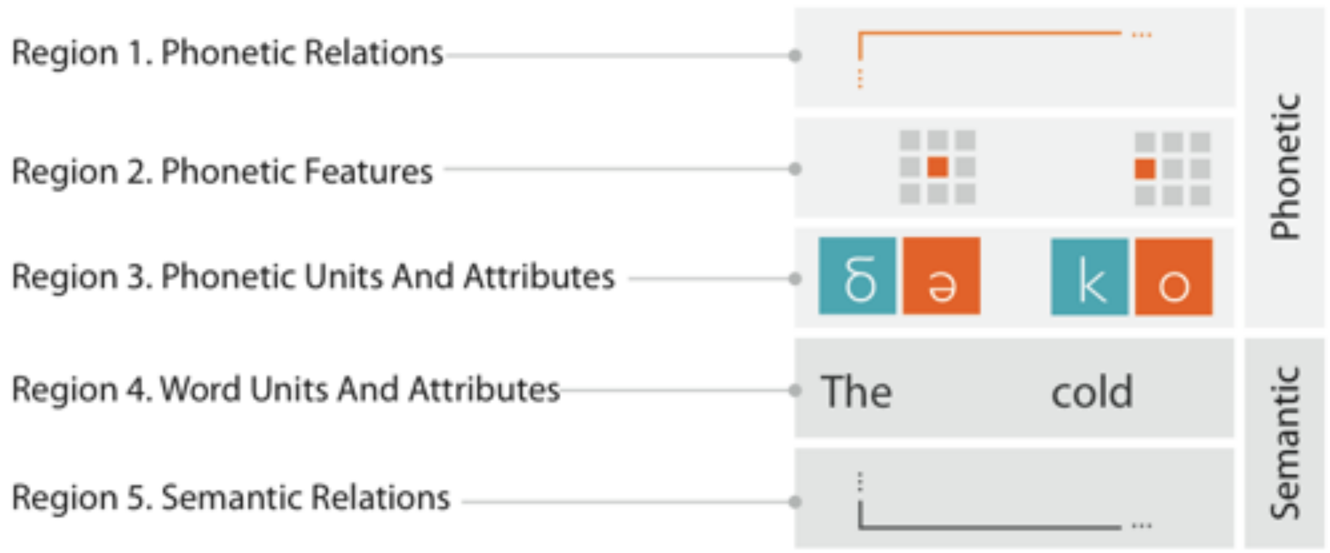


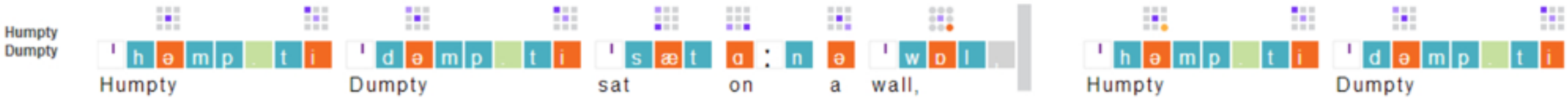


Analysis of the Characters from Les Misérables: http://mbostock.github.io/protovis/ex/arc.html

Rule-Based: Poetry

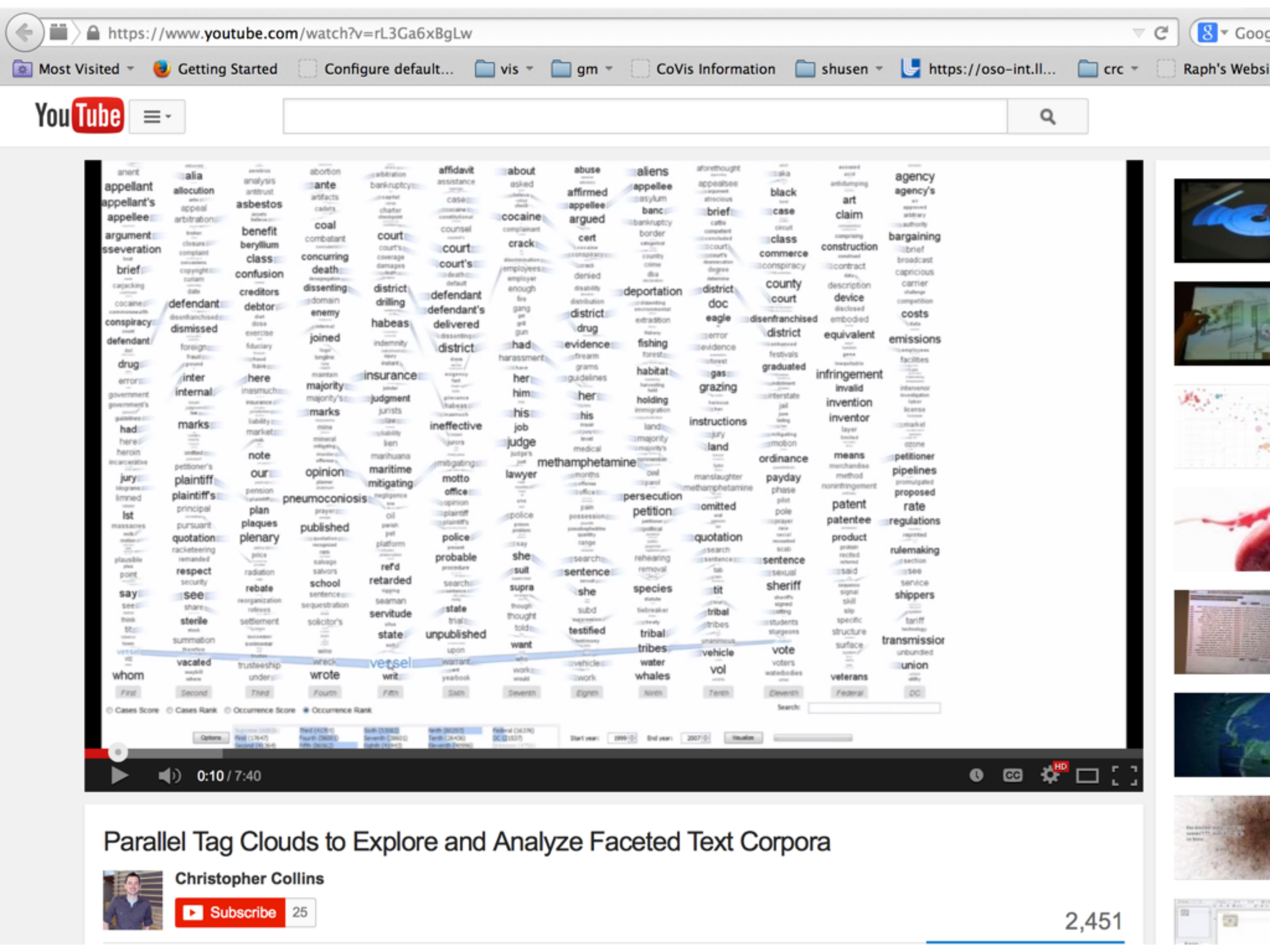



Region 1. Phonetic Relations
Region 2. Phonetic Features
Region 3. Phonetic Units And Attributes
Region 4. Word Units And Attributes
Region 5. Semantic Relations
ð
ə
k
o
The
cold
Phonetic
Semantic



Abdul-Rahman et al. 2011

collection of documents



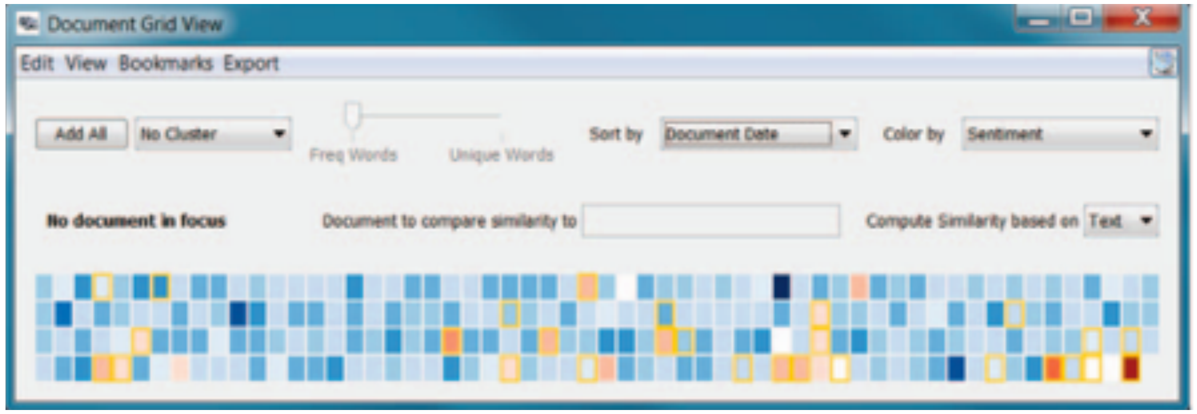
Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora

Christopher Collins

Subscribe 25

2,451

Document Cards
(small multiples)


Document Grid View

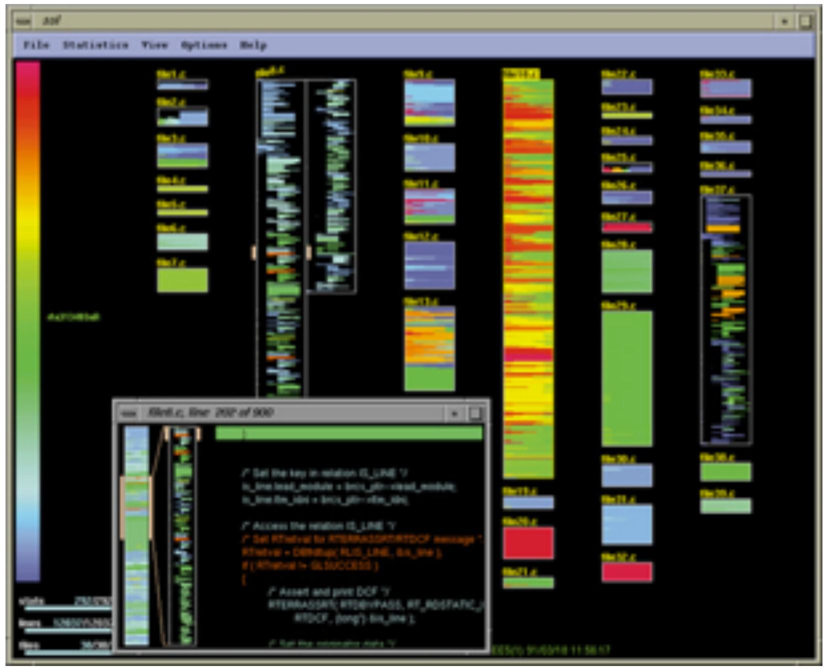
Edit View Bookmarks Export

Add All | No Cluster | Freq Words | Unique Words | Sort by: Document Date | Color by: Sentiment

No document in focus

Document to compare similarity to:

Compute Similarity based on: Text

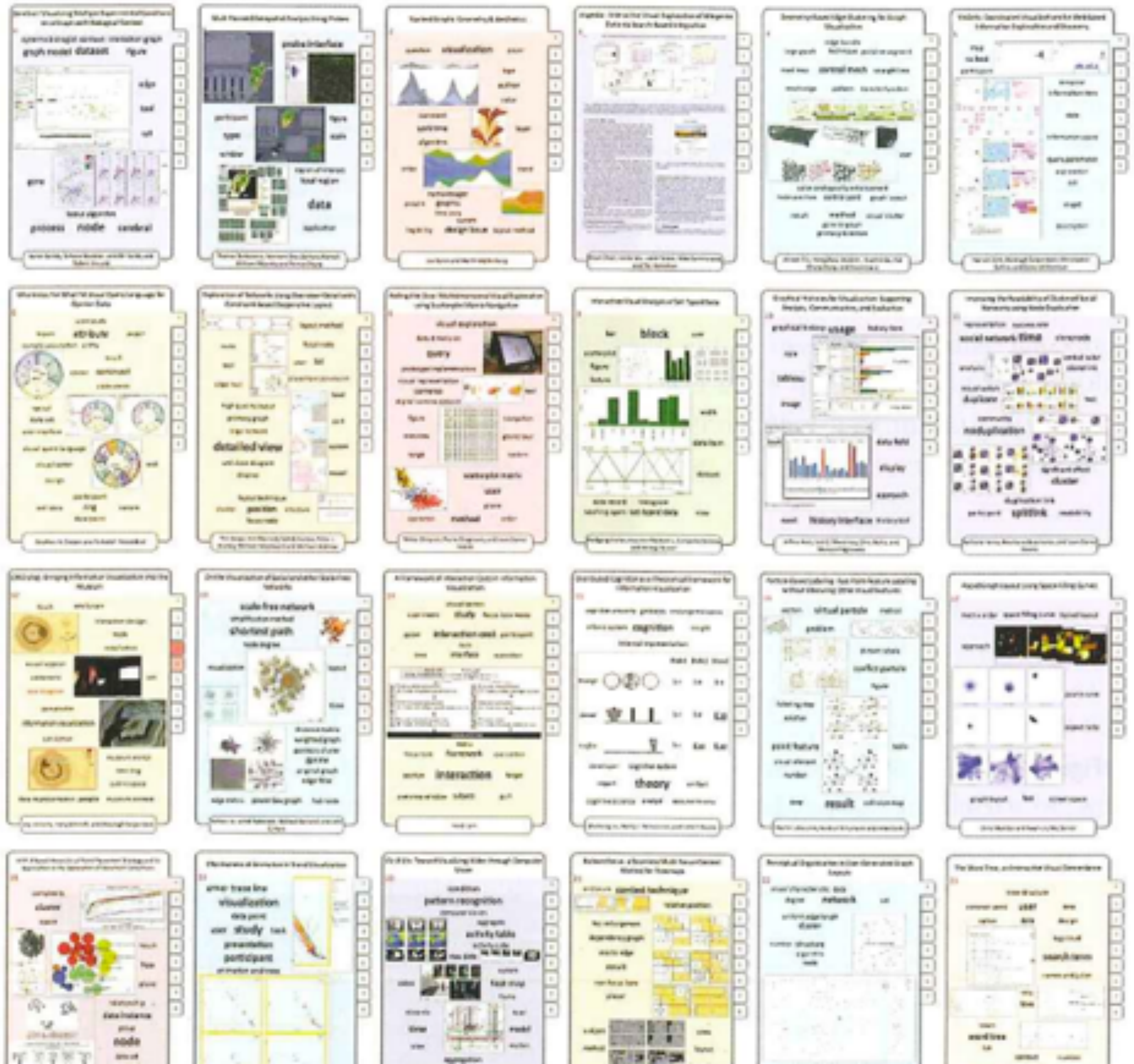


file8.c, line 202 of 900

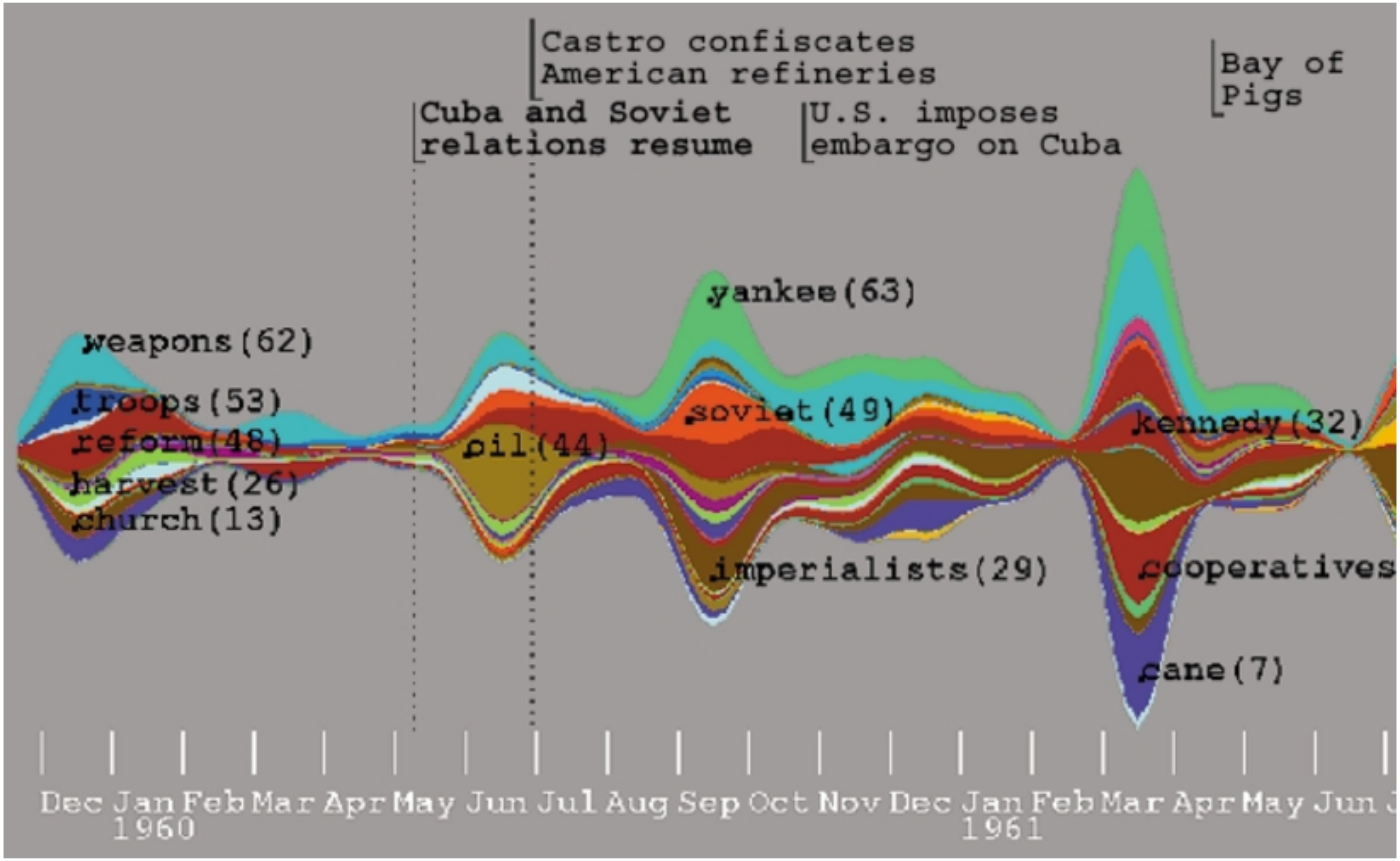
```
/* Set the key in relation IS_LINE */
is_line.load_module = bncs_pln->load_module;
is_line.line_sbi = bncs_pln->line_sbi;

/* Access the relation IS_LINE */
/* Set RTretval for RTERRASSRTDCF message */
RTretval = DBfindtup( RLIS_LINE, &is_line );
if ( RTretval != GLSUCCESS )
{
    /* Assert and print DCF */
    RTERRASSRT( RTDBYPASS, RT_RDSTATIC_
        RTDCF, (long) &is_line );

    /* Set the [illegible] */
```

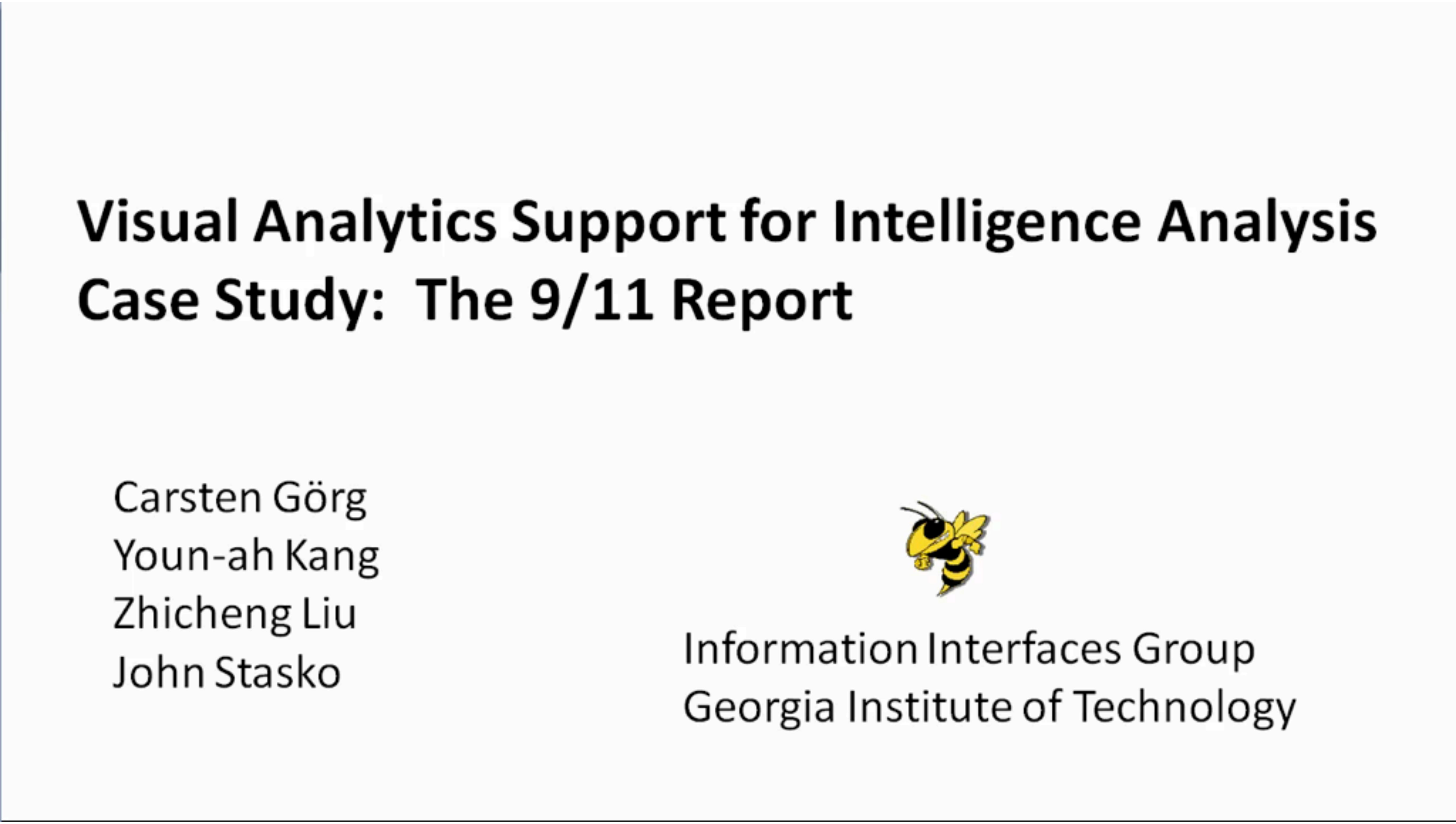


Showing Temporal Relationships: ThemeRiver (Stream Graph)



Havre, Hetzler, Nowell 2000

Jigsaw: Many Linked Views



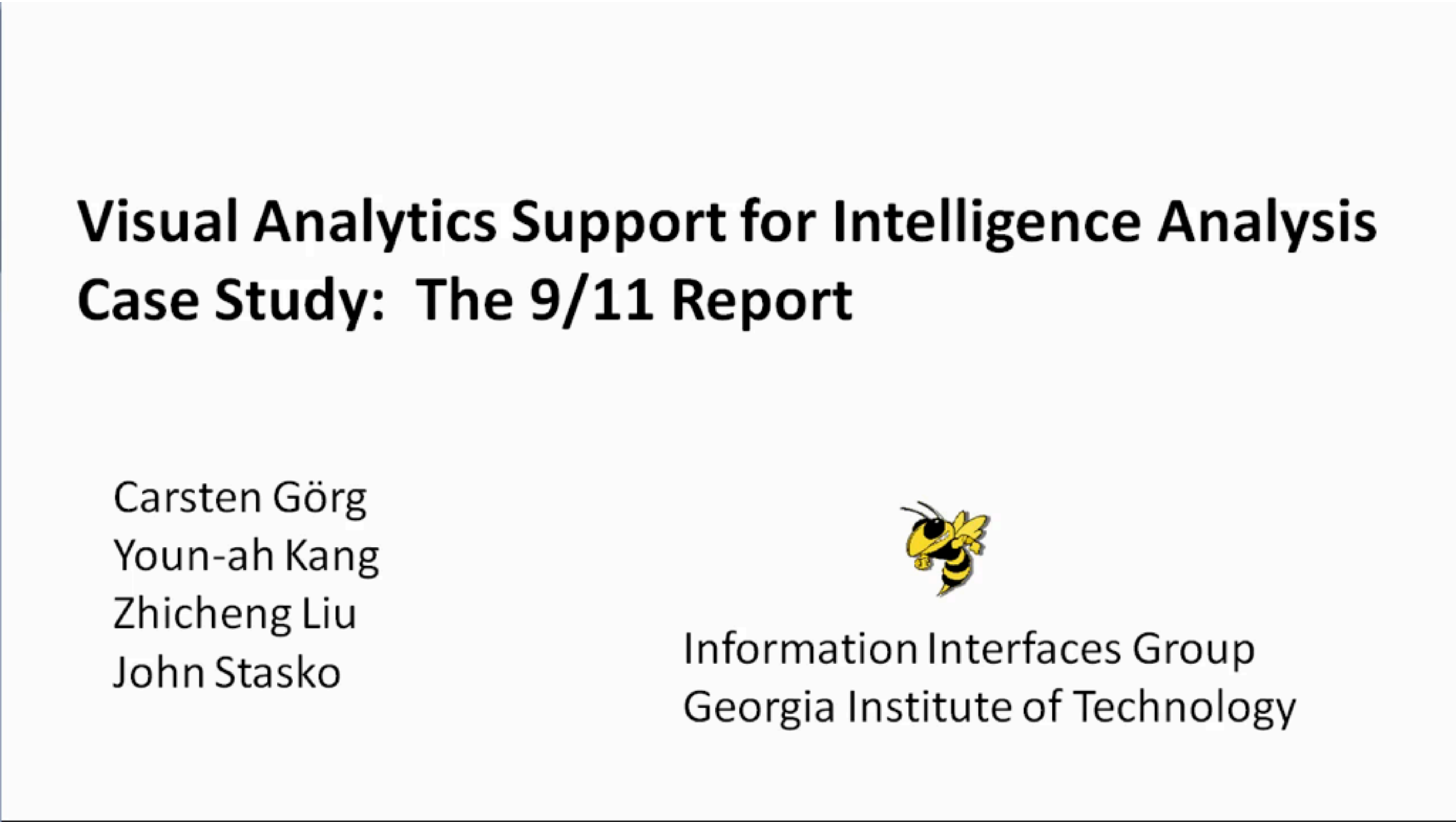
Visual Analytics Support for Intelligence Analysis
Case Study: The 9/11 Report

Carsten Görg
Youn-ah Kang
Zhicheng Liu
John Stasko

Information Interfaces Group
Georgia Institute of Technology

Stasko et al. 2008

Jigsaw: Many Linked Views



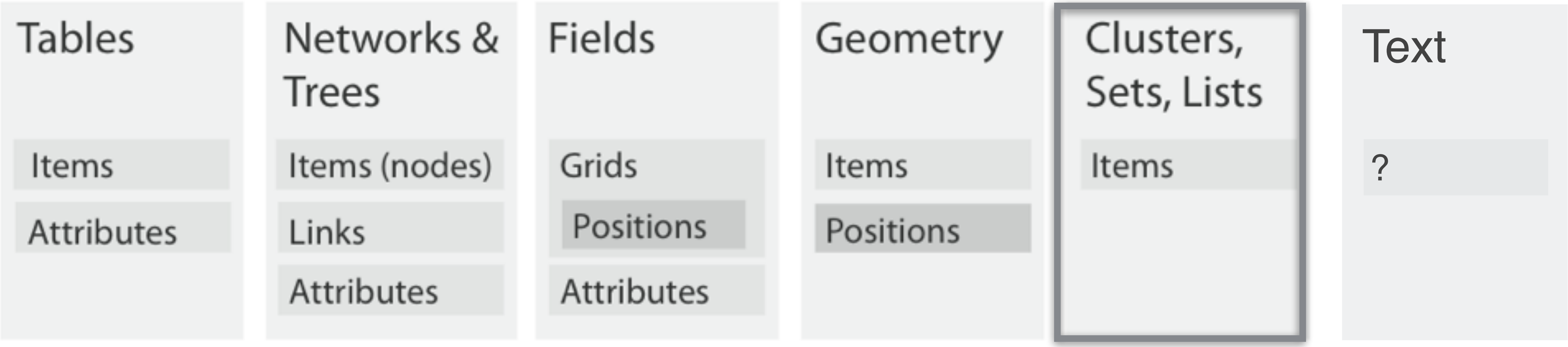
Visual Analytics Support for Intelligence Analysis
Case Study: The 9/11 Report

Carsten Görg
Youn-ah Kang
Zhicheng Liu
John Stasko

Information Interfaces Group
Georgia Institute of Technology

Stasko et al. 2008

SETS



Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

Geometry

Items

Positions

Clusters, Sets, Lists

Items

Text

?

thought experiment...

- item: Lego

- attributes

thought experiment...

- **item:** Lego

- **attributes**
  - color
  - height
  - width
  - length
  - shape

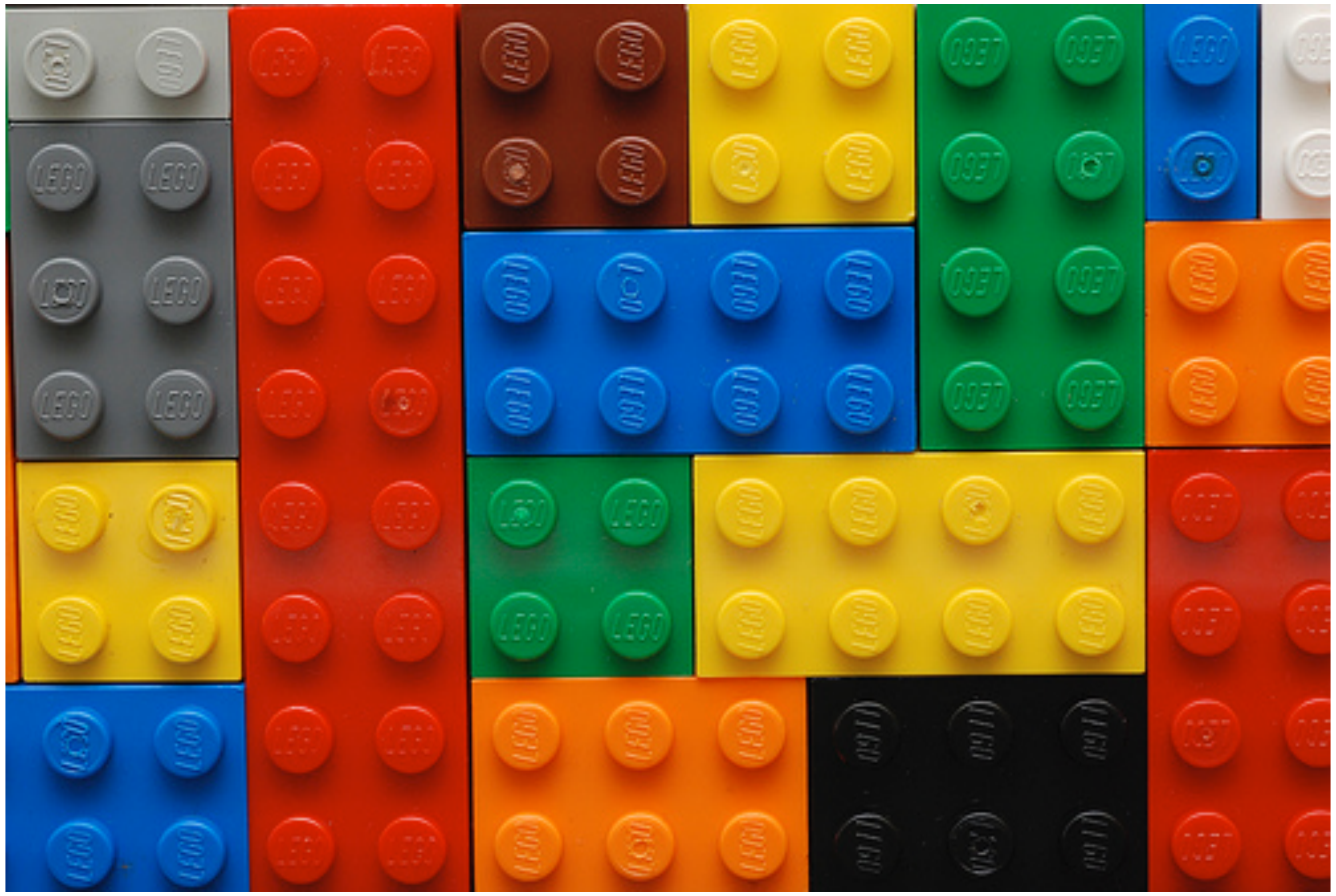
dataset: option 1

dataset: option 2

dataset: more realistic

dataset

- **where do we start?** we need to organize! but, how?

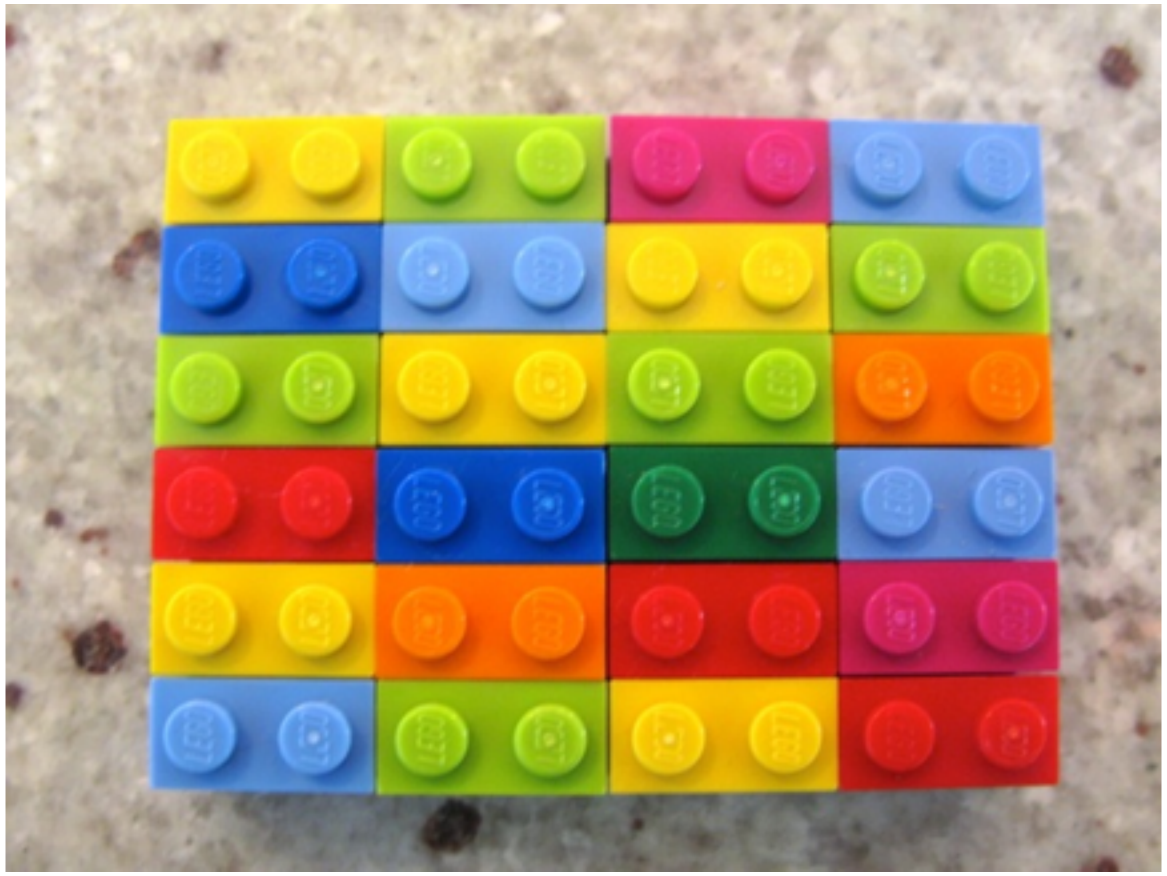


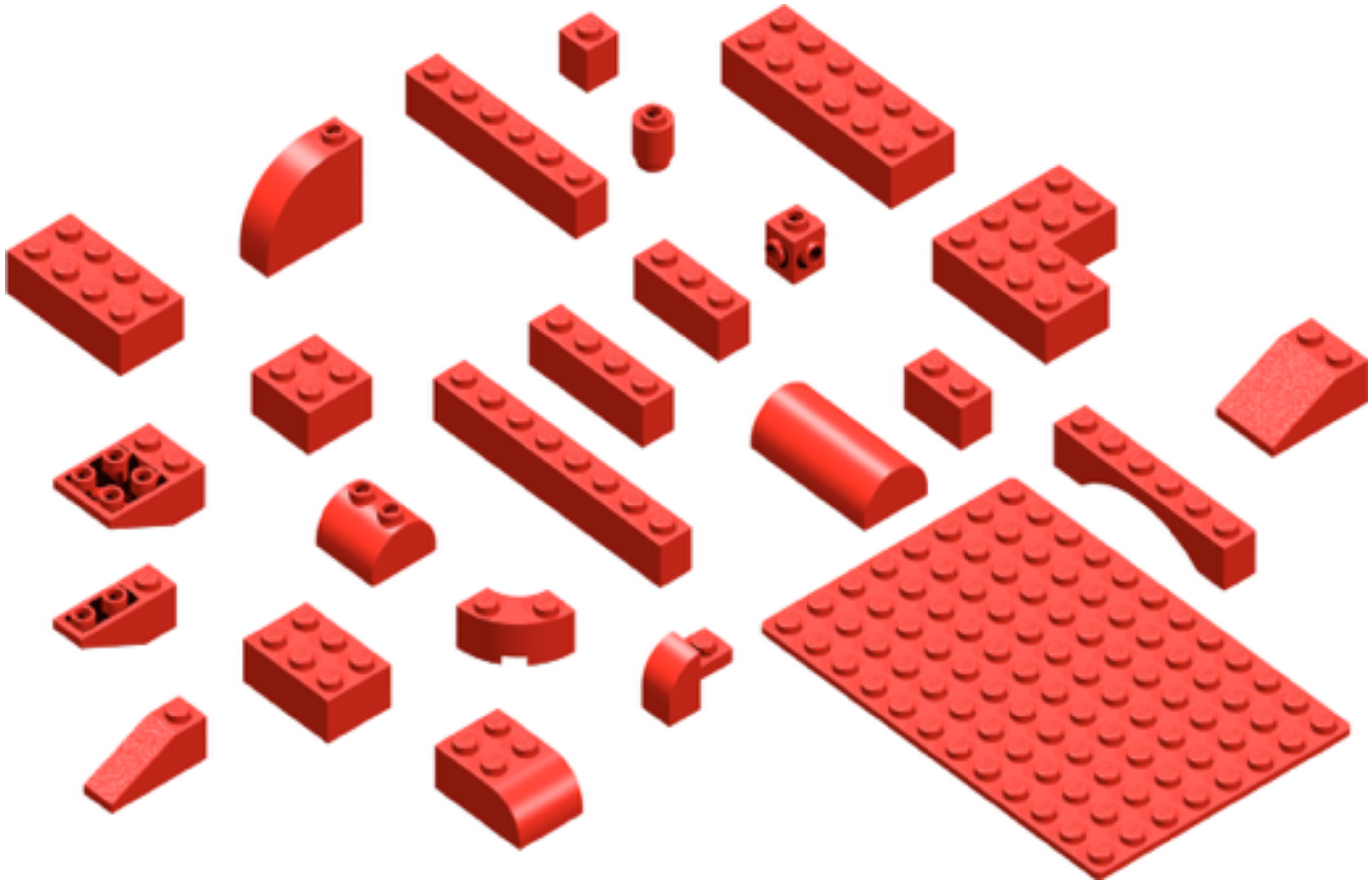
dataset

- sort by color

dataset

- sort by size, shape





dataset

- **task:** organization

- drawbacks?

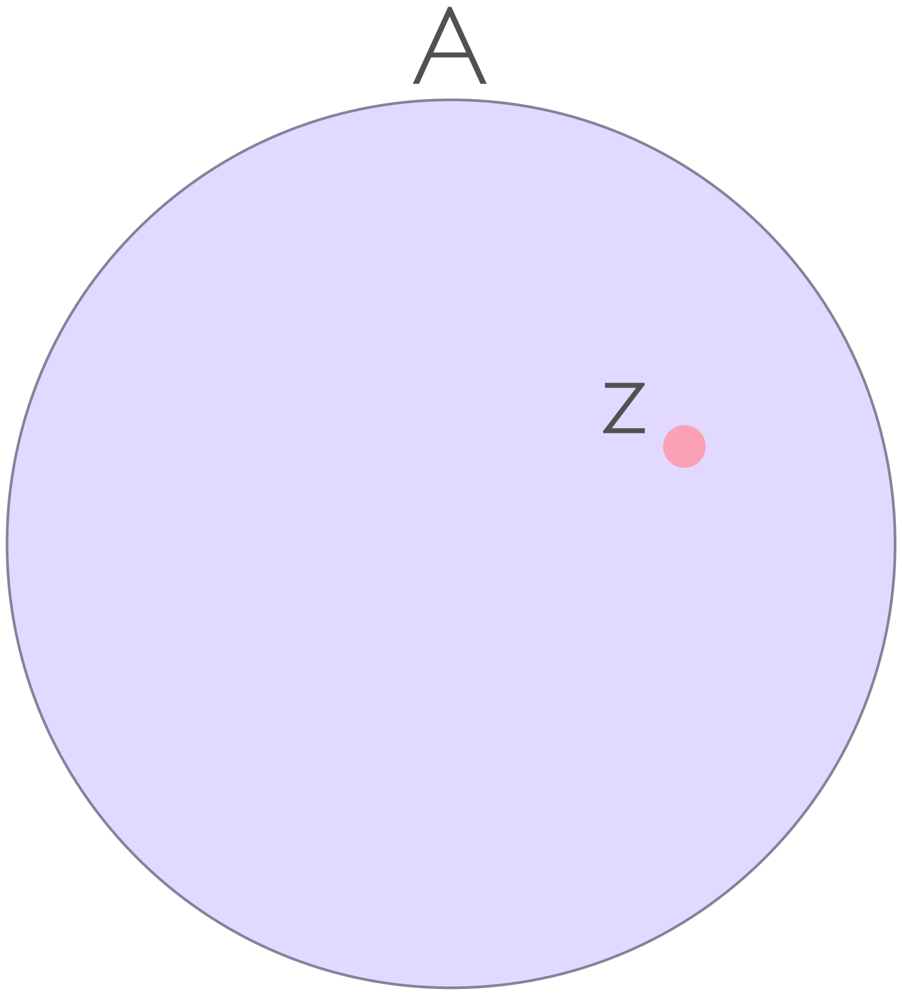
dataset

- organization leads us to a set problem

- so what are sets?

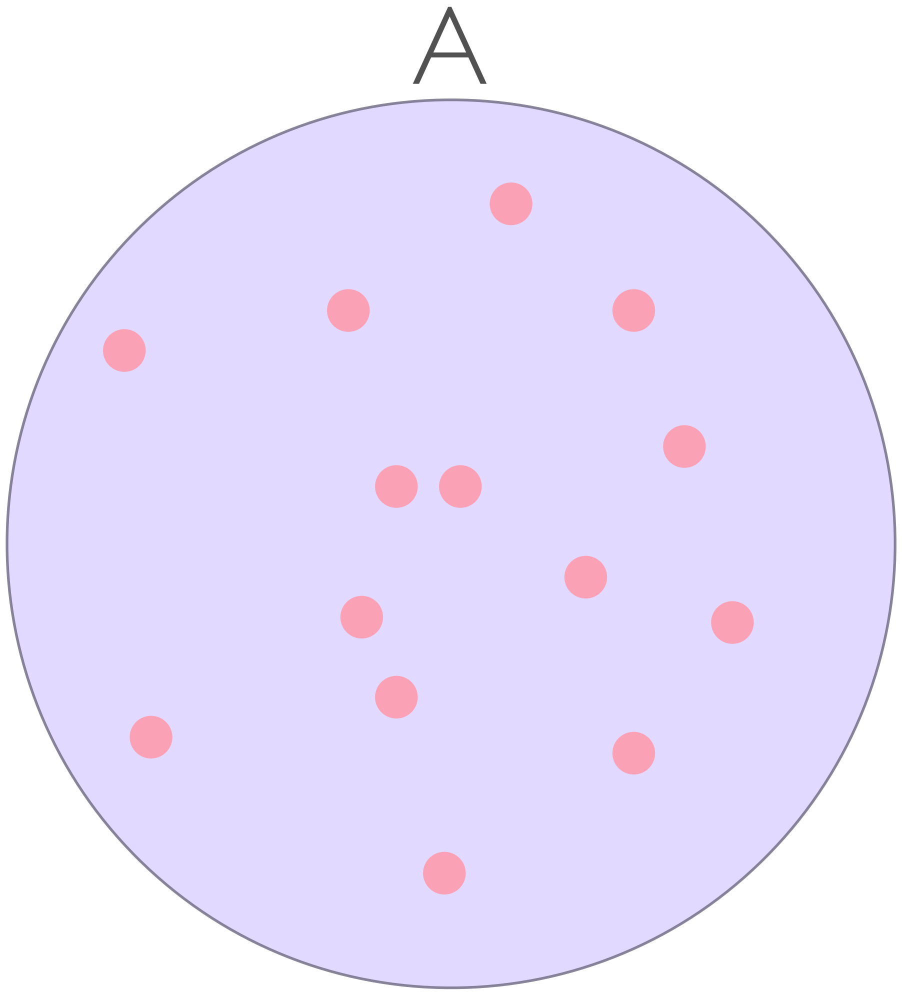
set theory

- **set**
  - a collection of objects
  - some set: A
- **object**
  - some object: z
  - $z \in A$



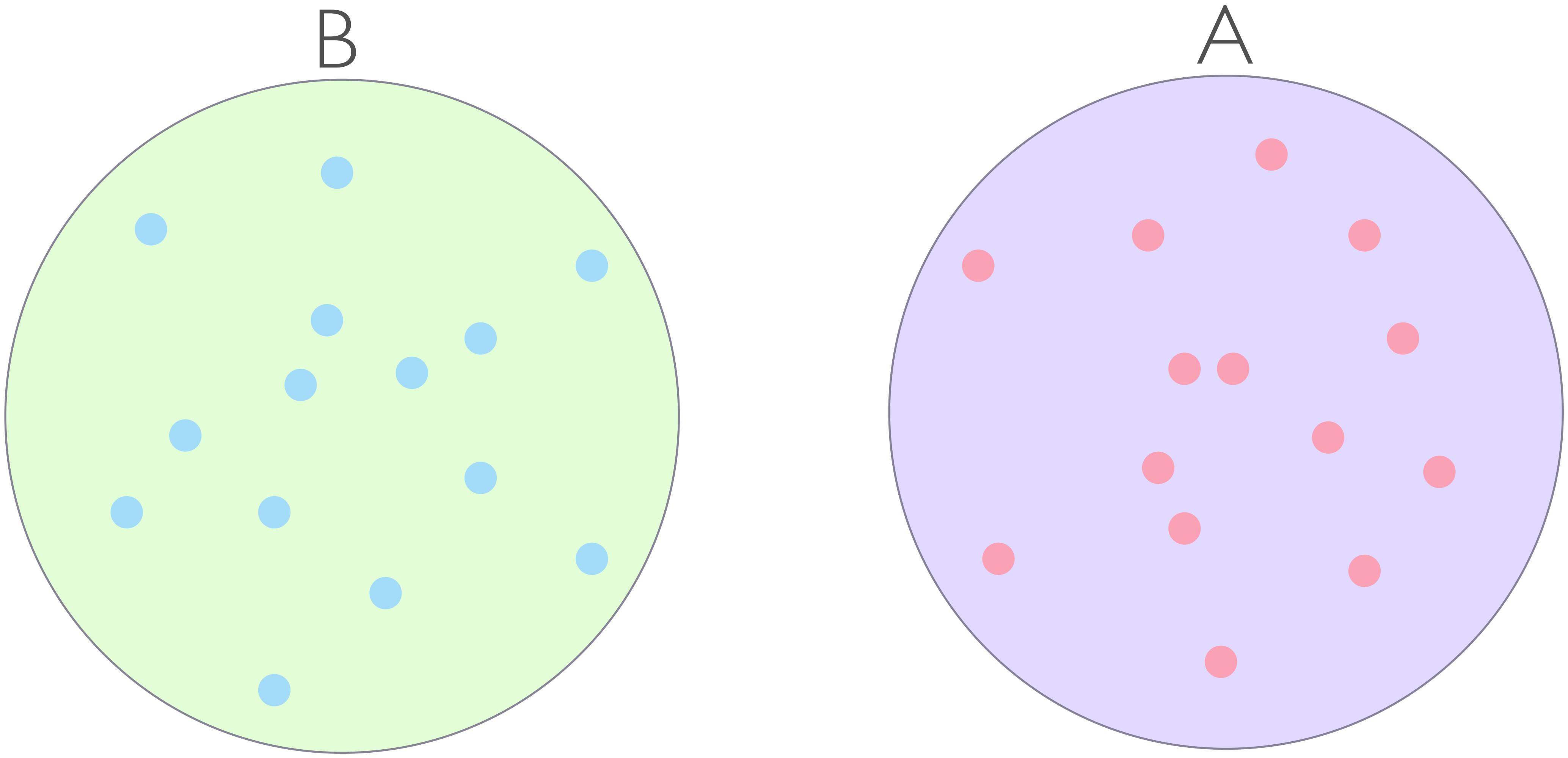
set theory

- **set**
  - a collection of objects
  - some set: A



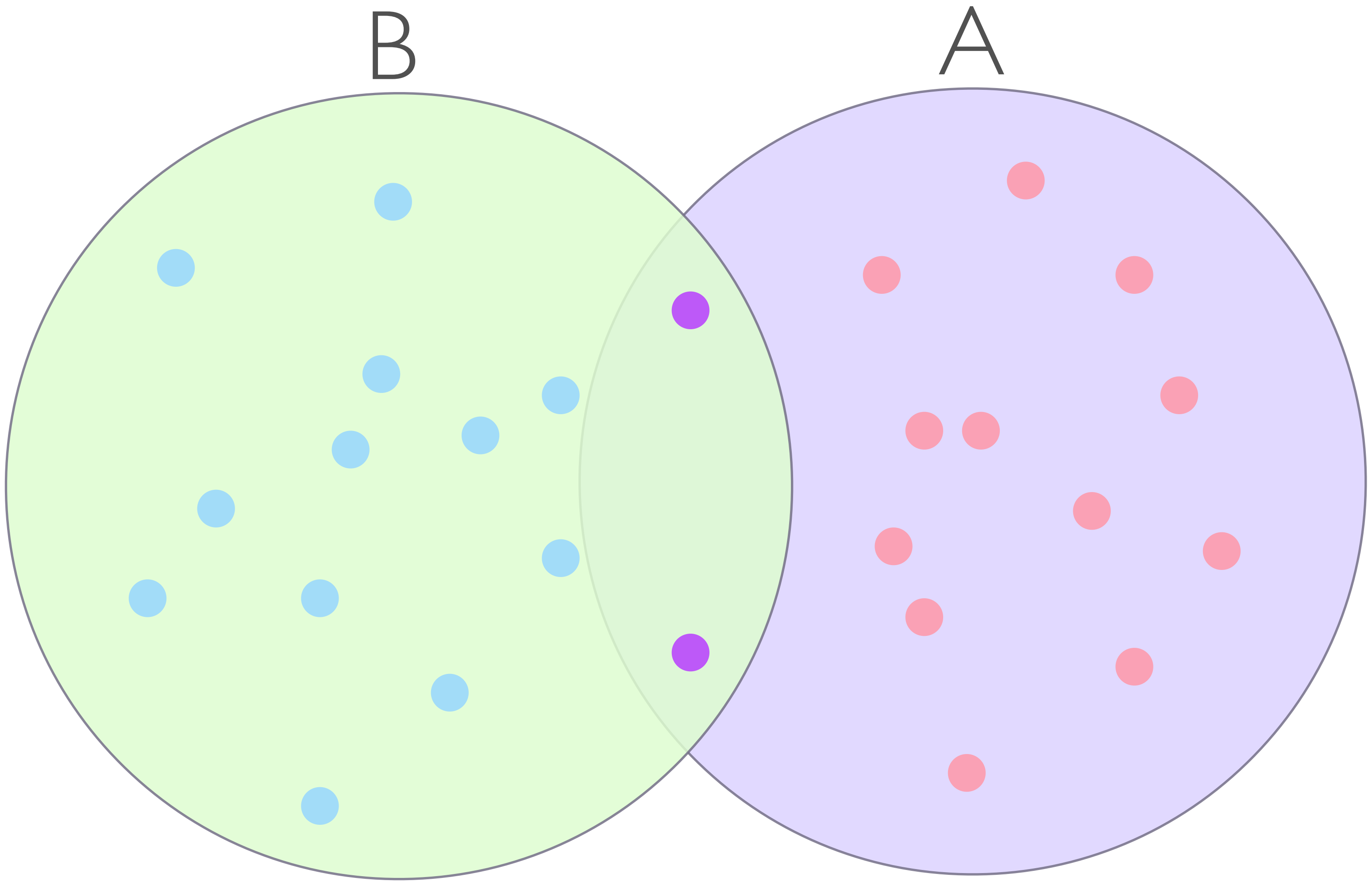
set theory

- multiple sets: A & B



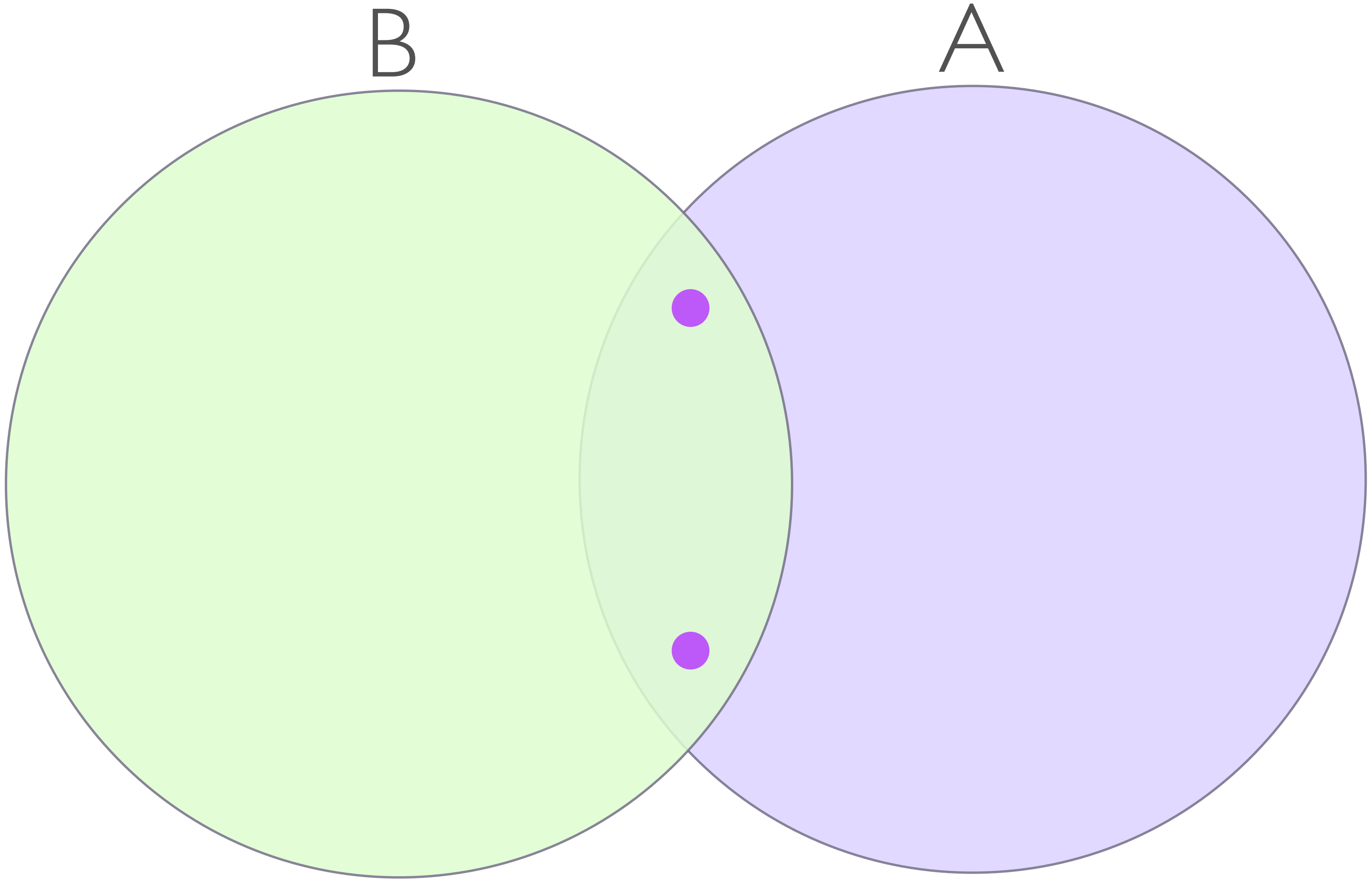
set theory

- union: A ∪ B



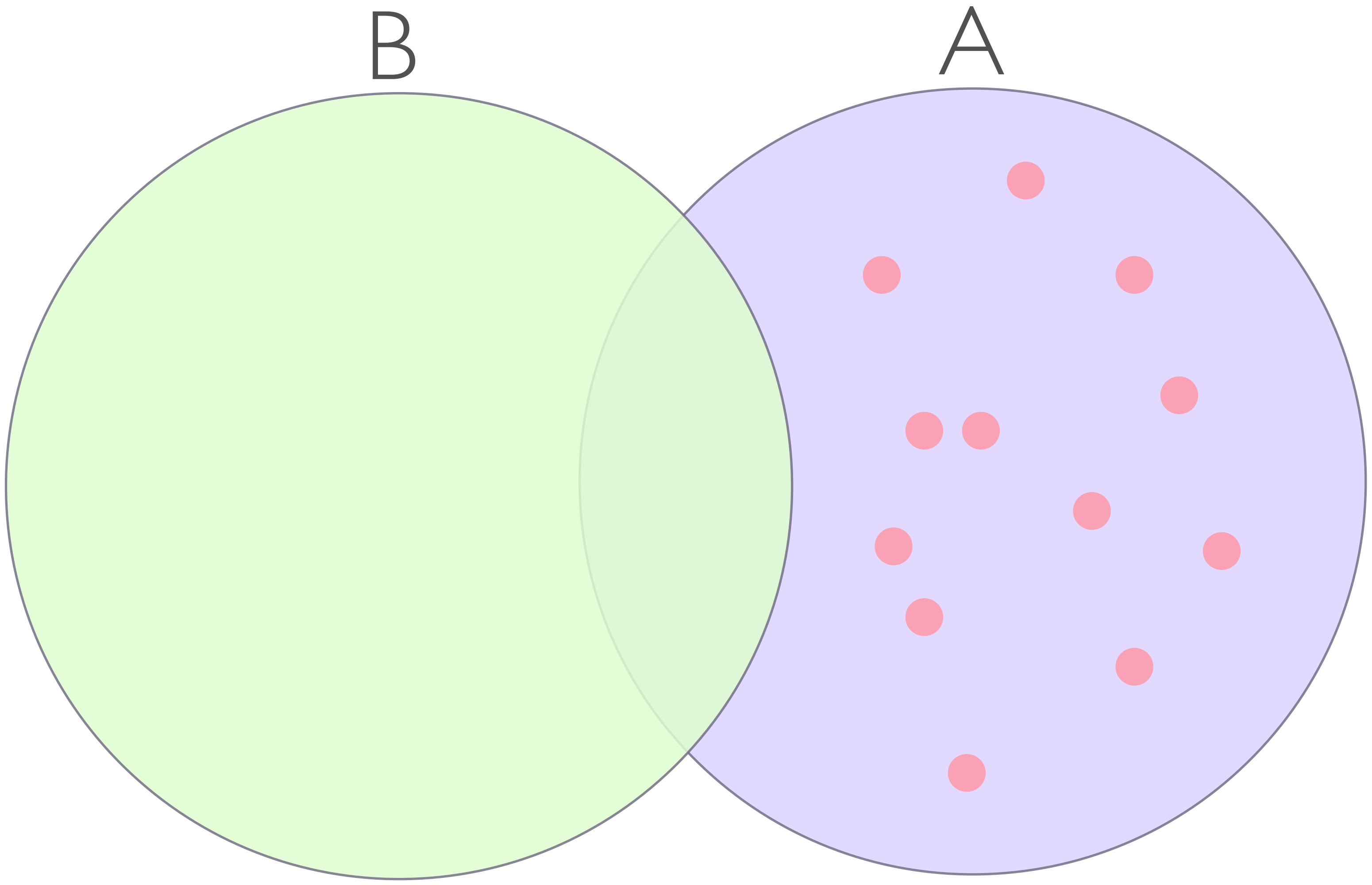
set theory

- intersection: $A \cap B$



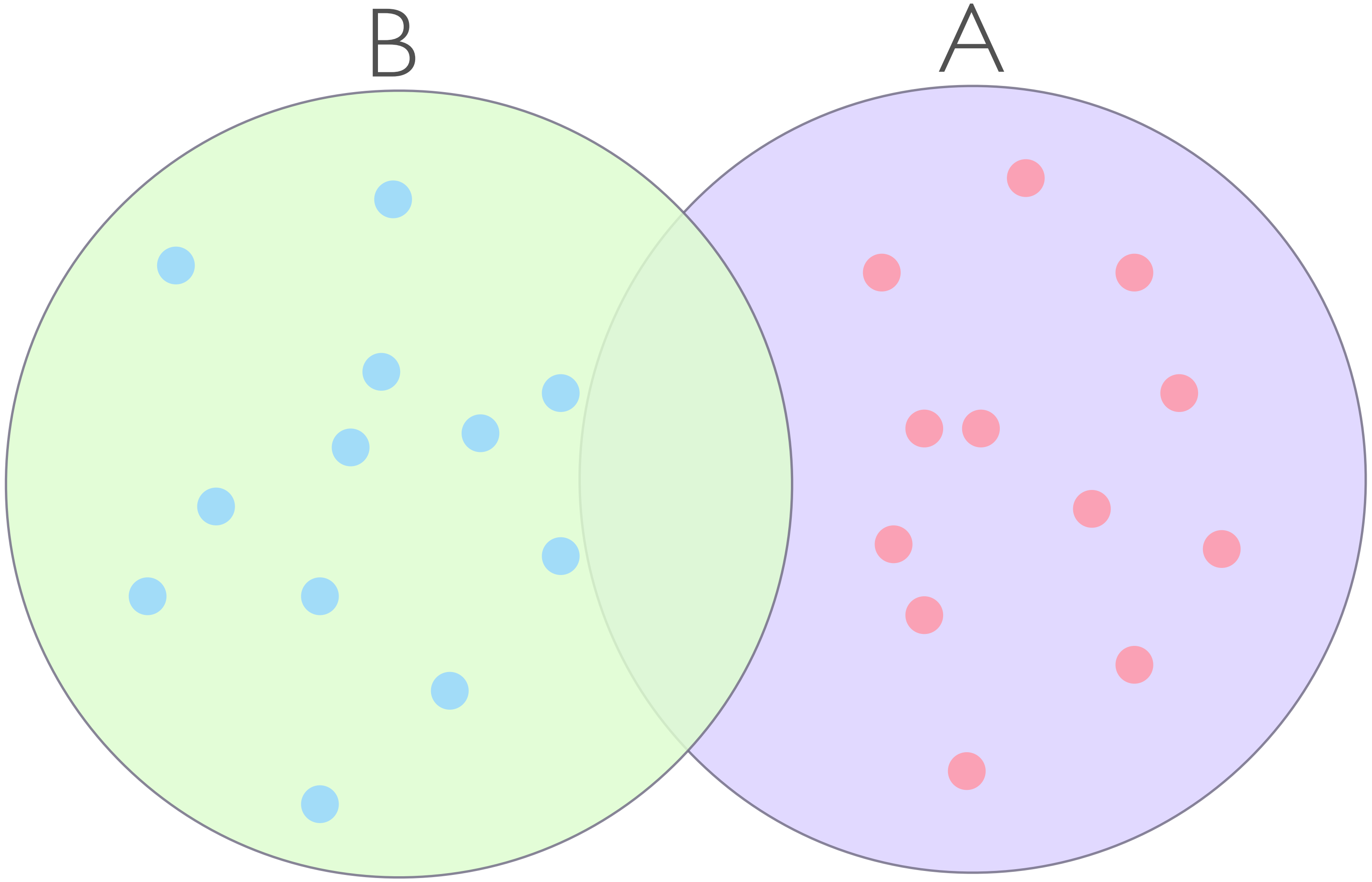
set theory

- set difference: $A \setminus B$



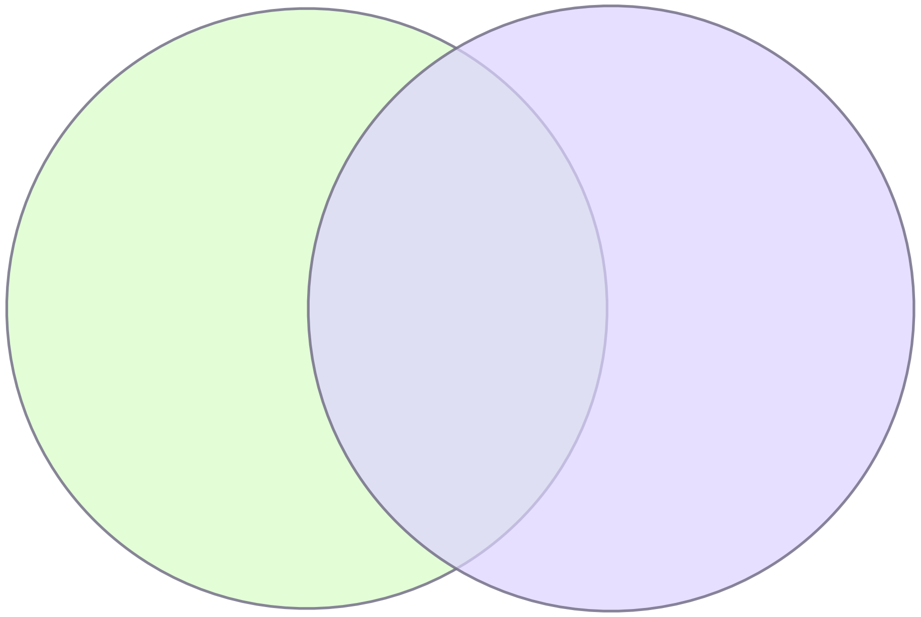
set theory

- symmetric difference: $A \ominus B$

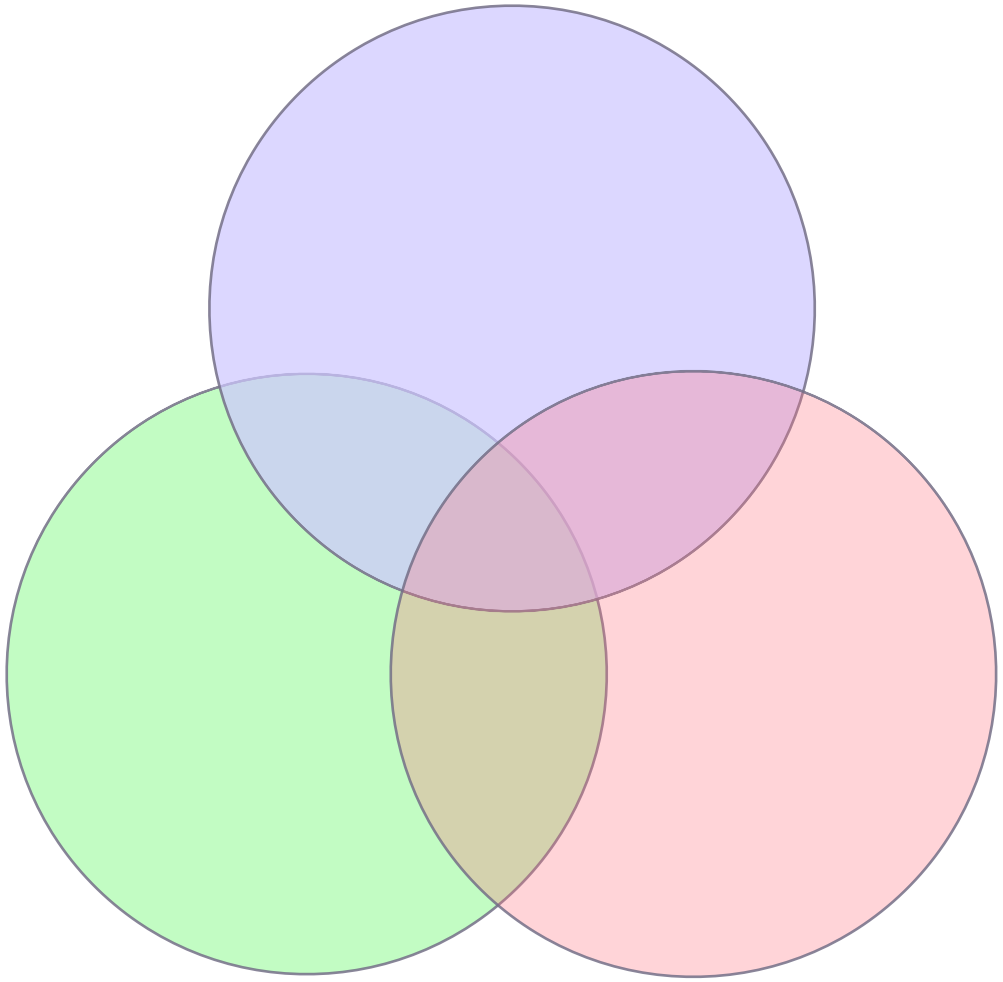


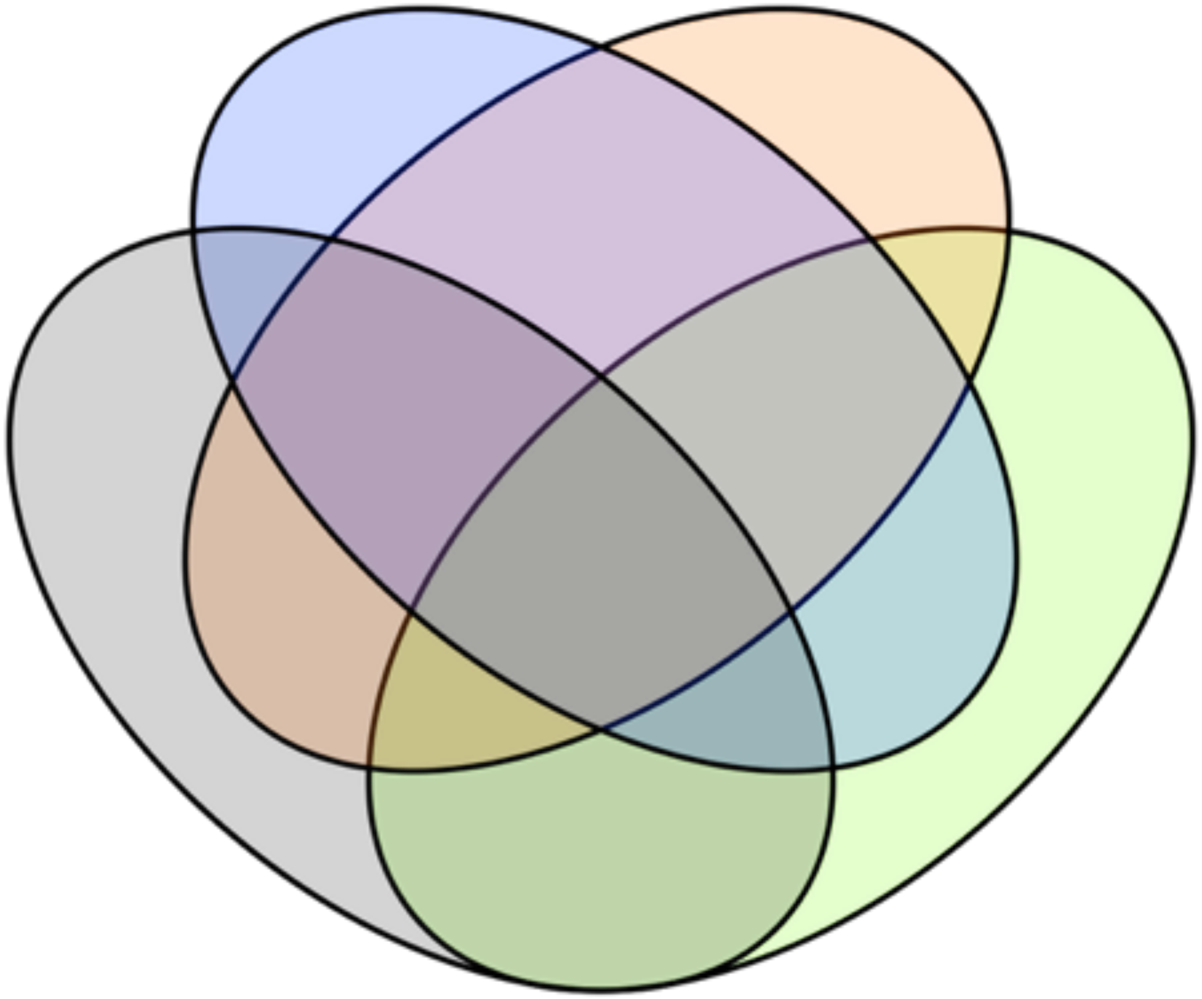
visualizing sets

venn diagrams



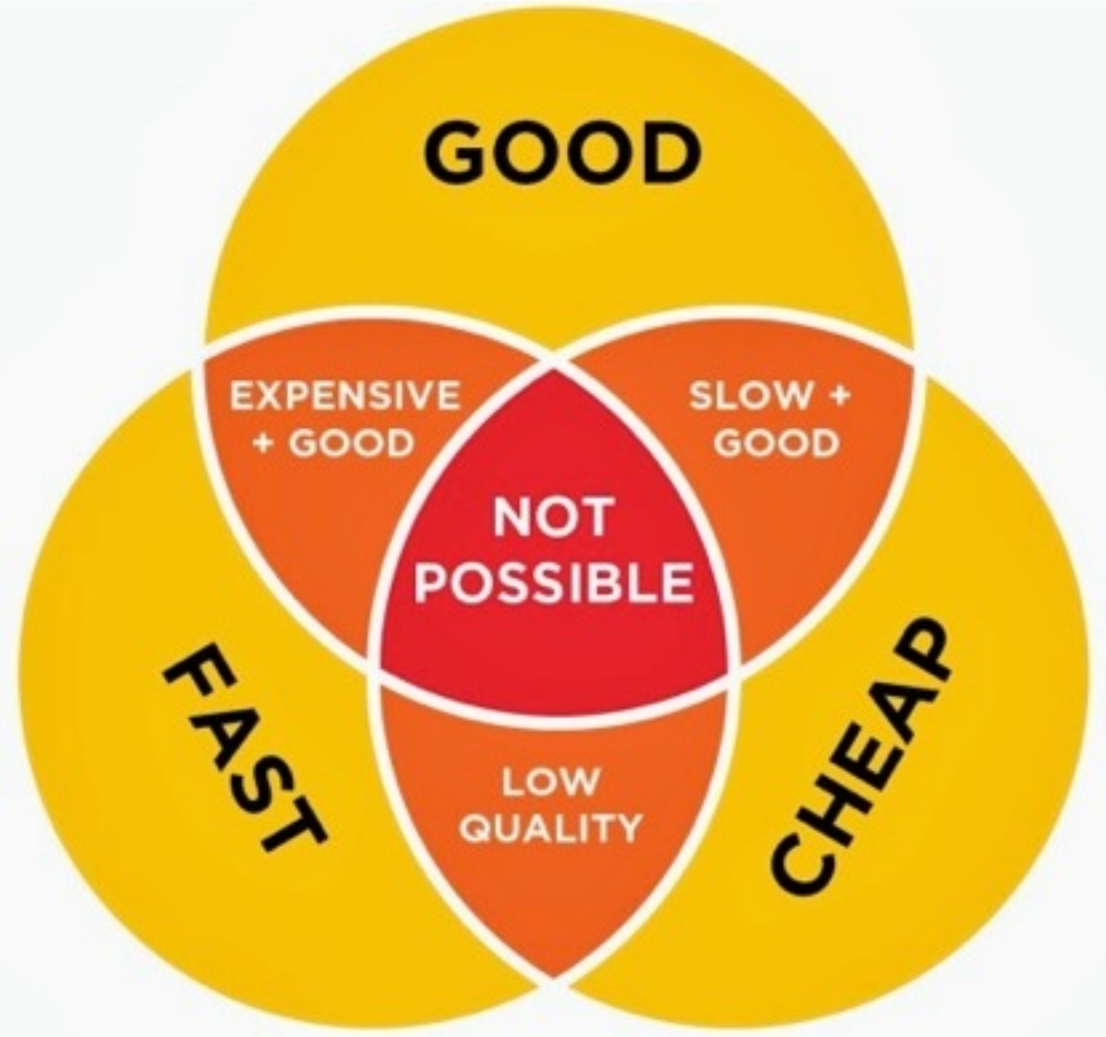
- show all possible relationships

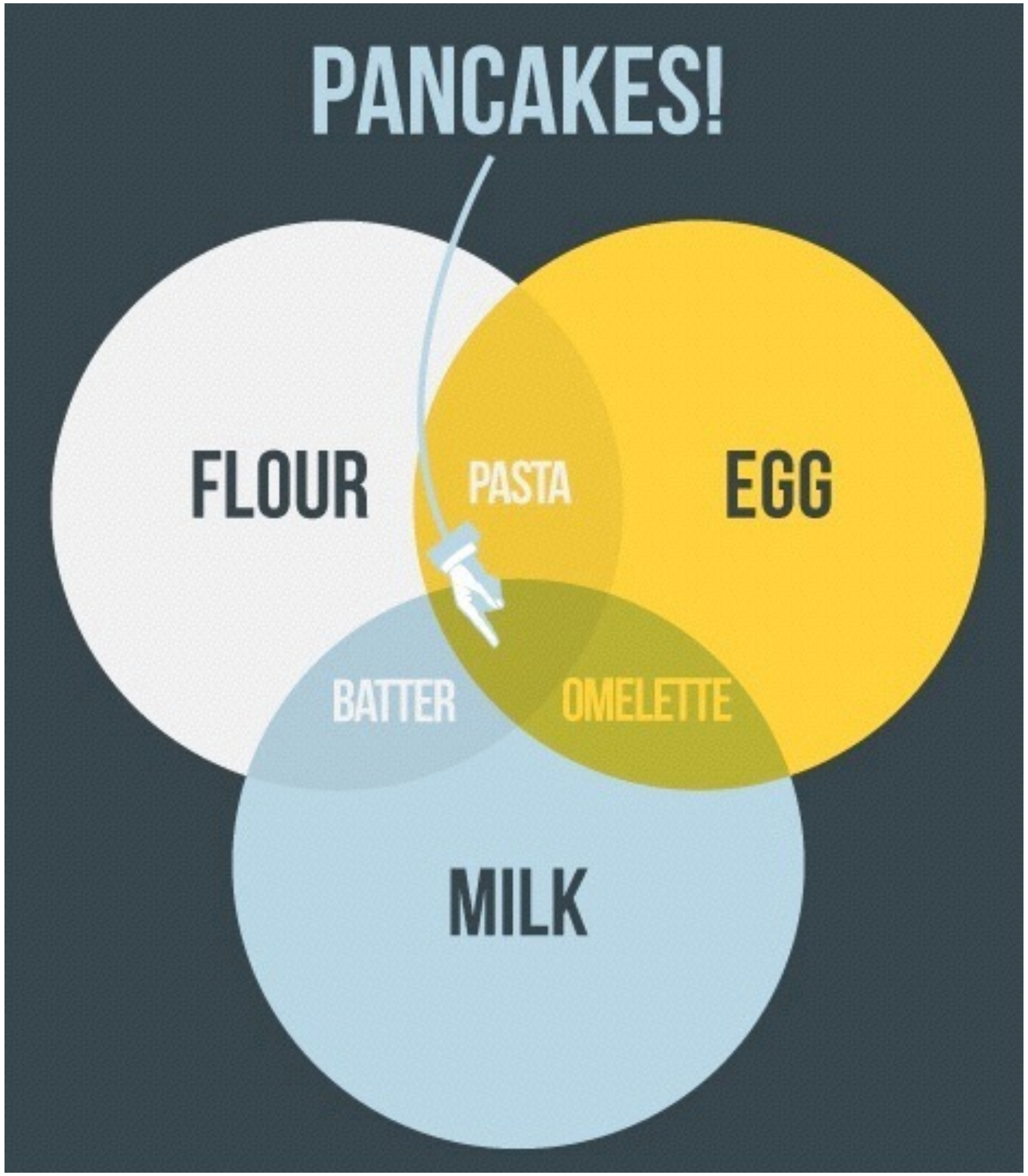




venn diagrams

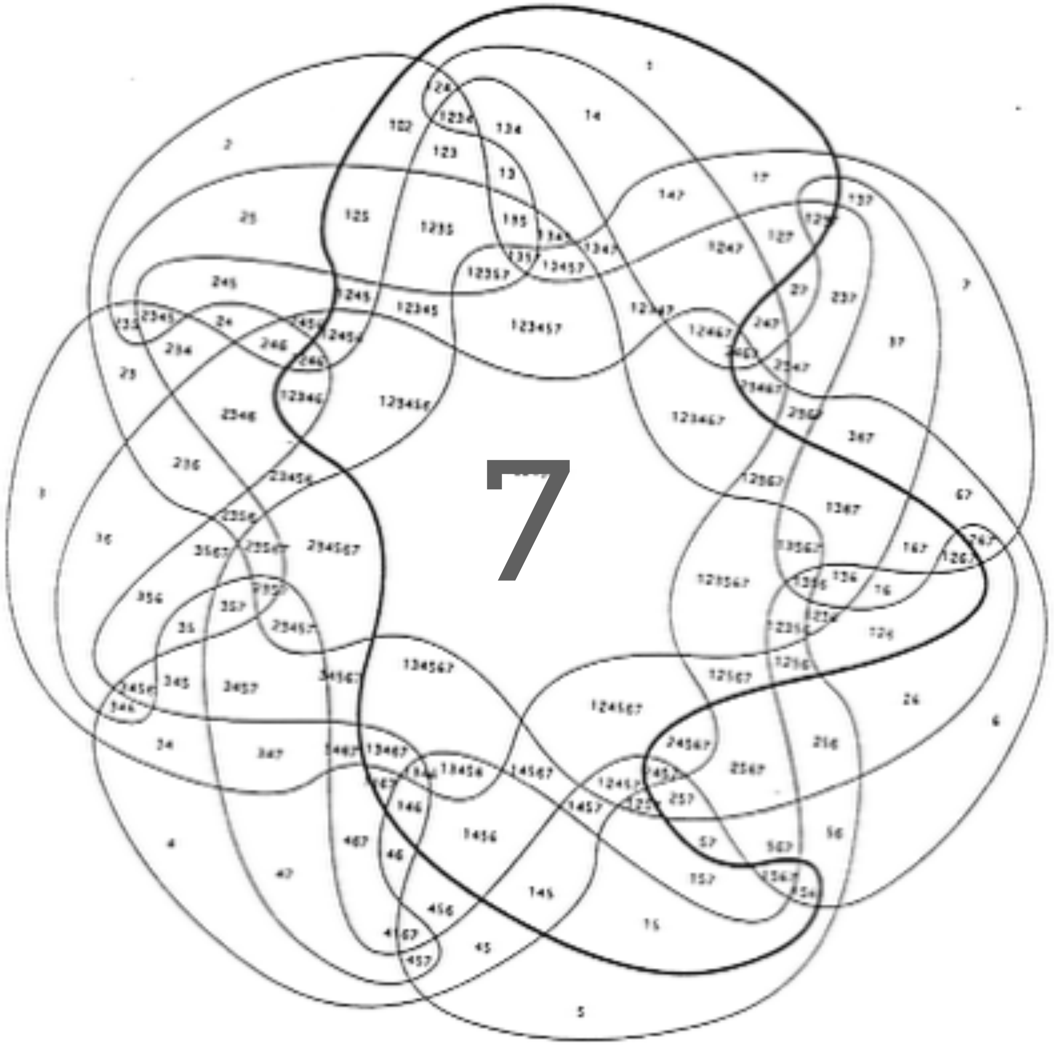
- casual infovis

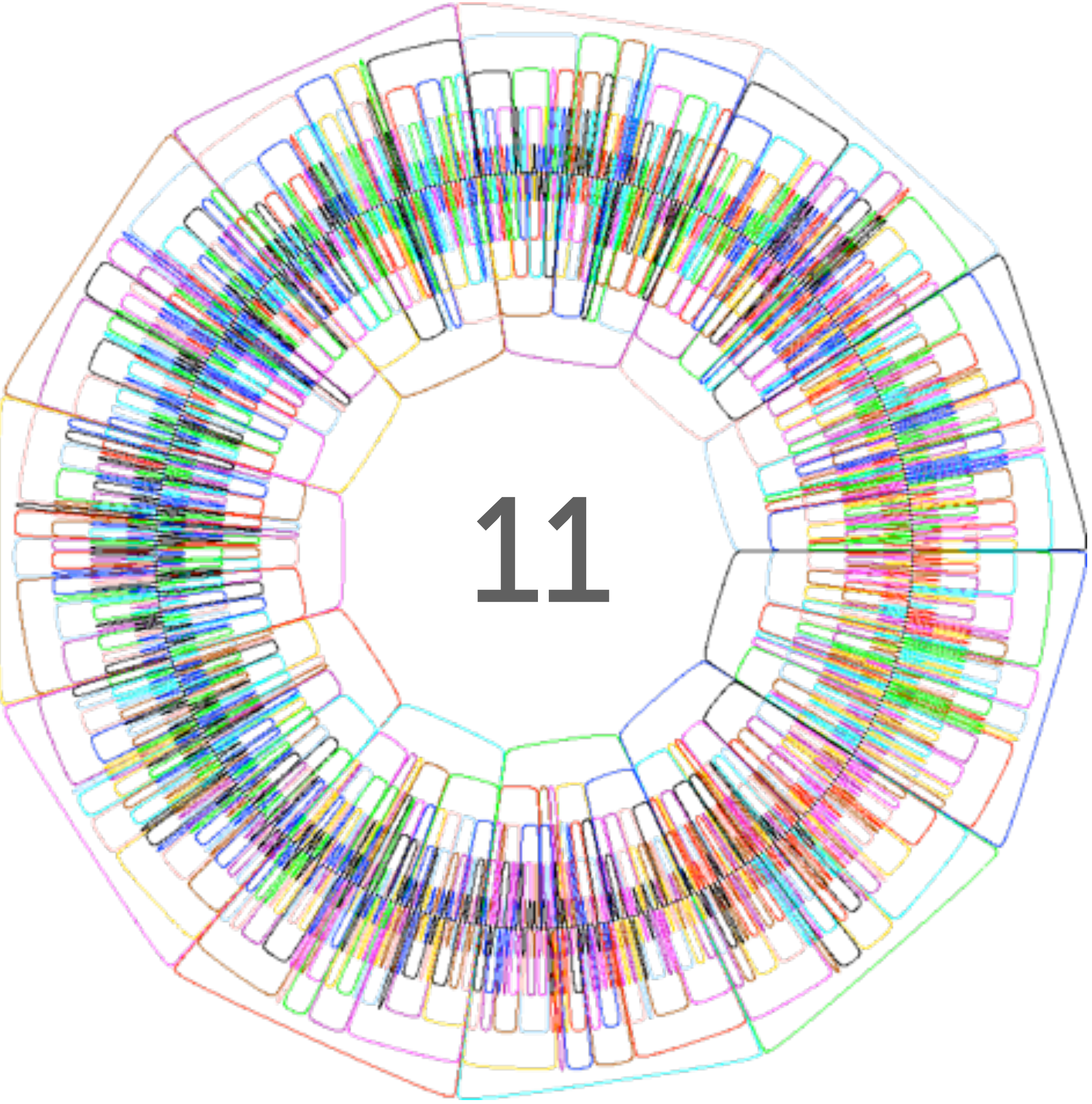




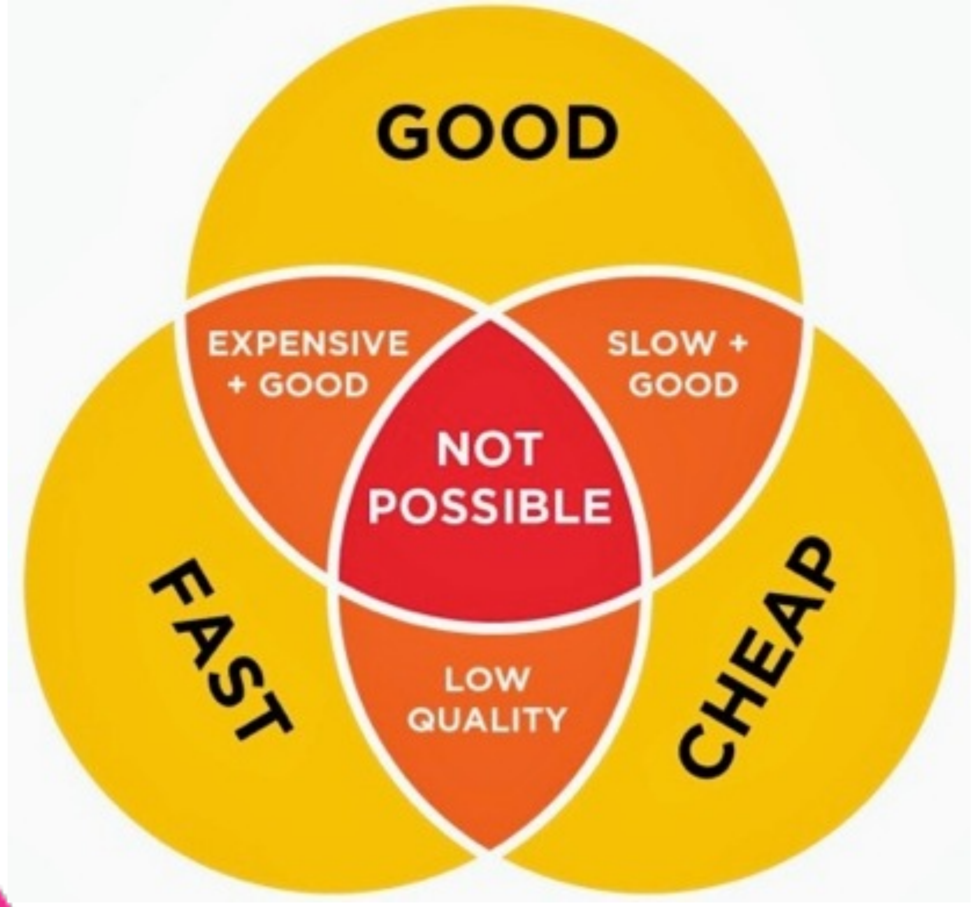
venn diagrams

- get messy fast

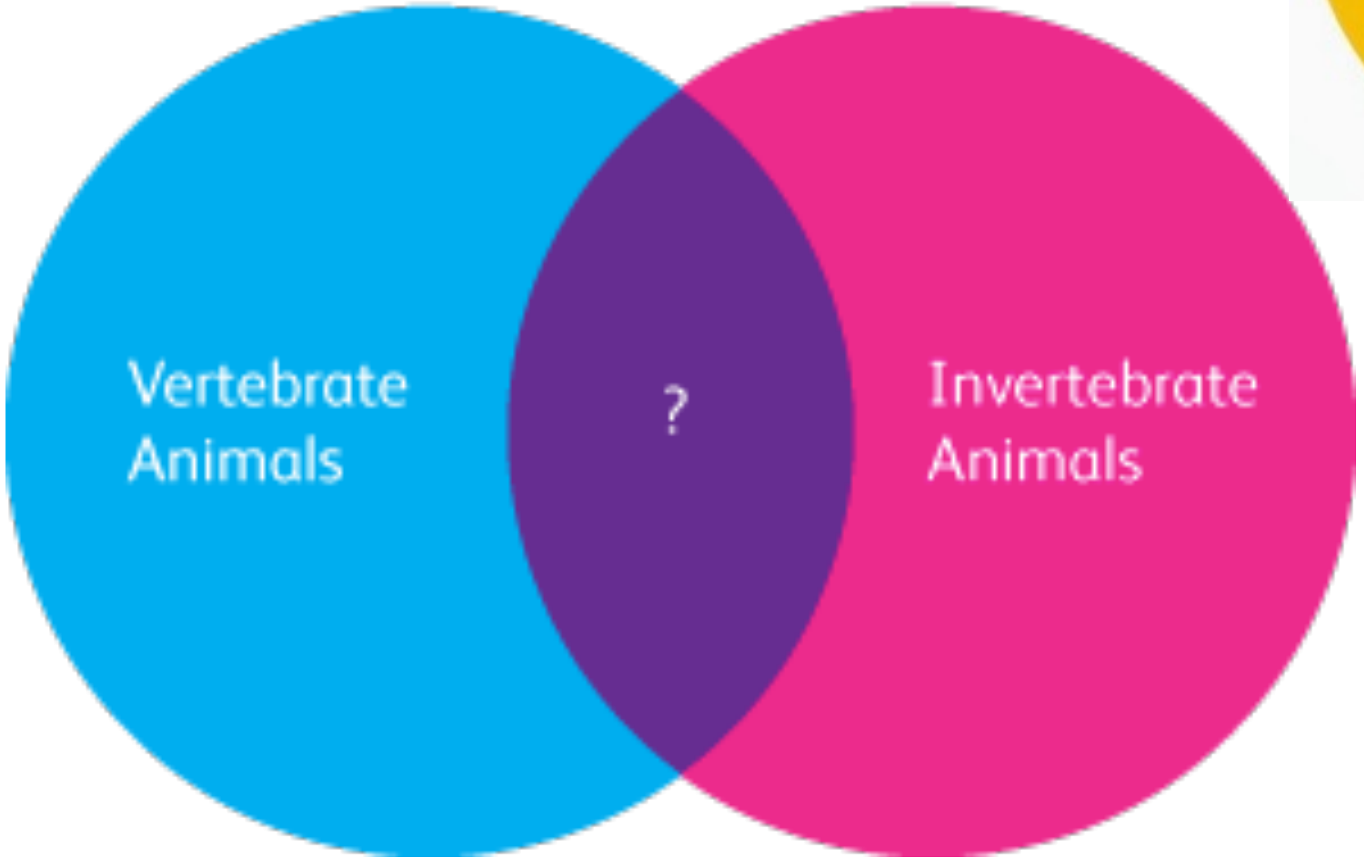




venn diagrams



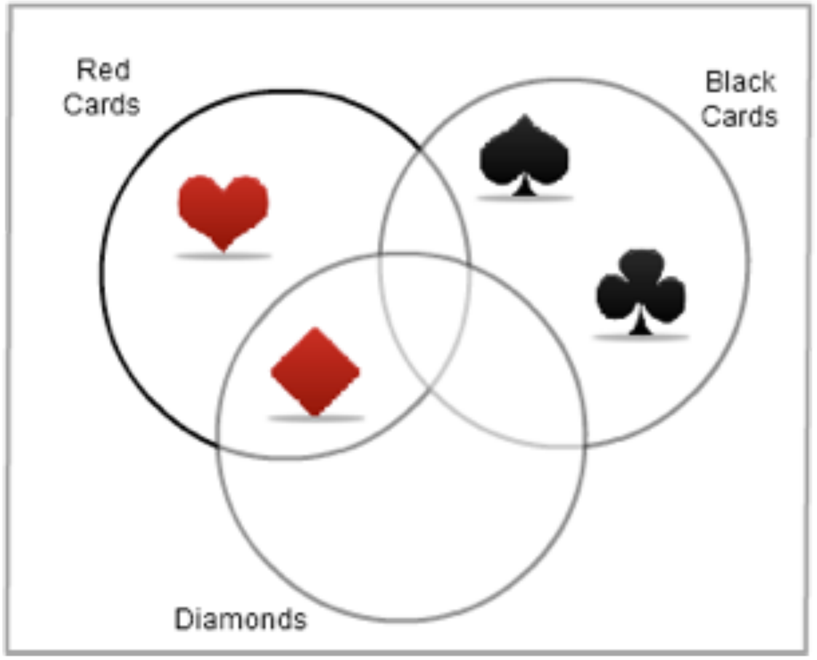
- non-sensical



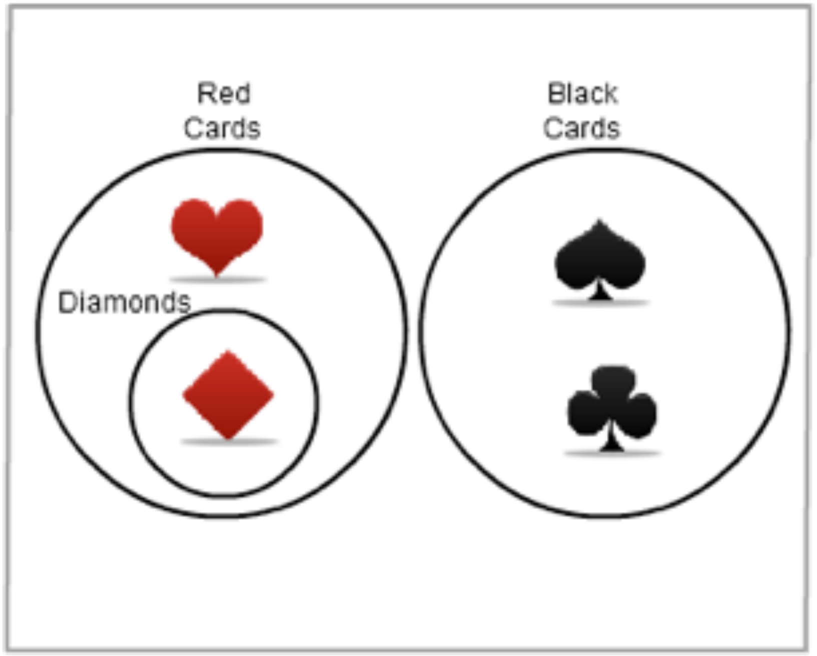
euler diagrams

- show only existing relationships

VENN

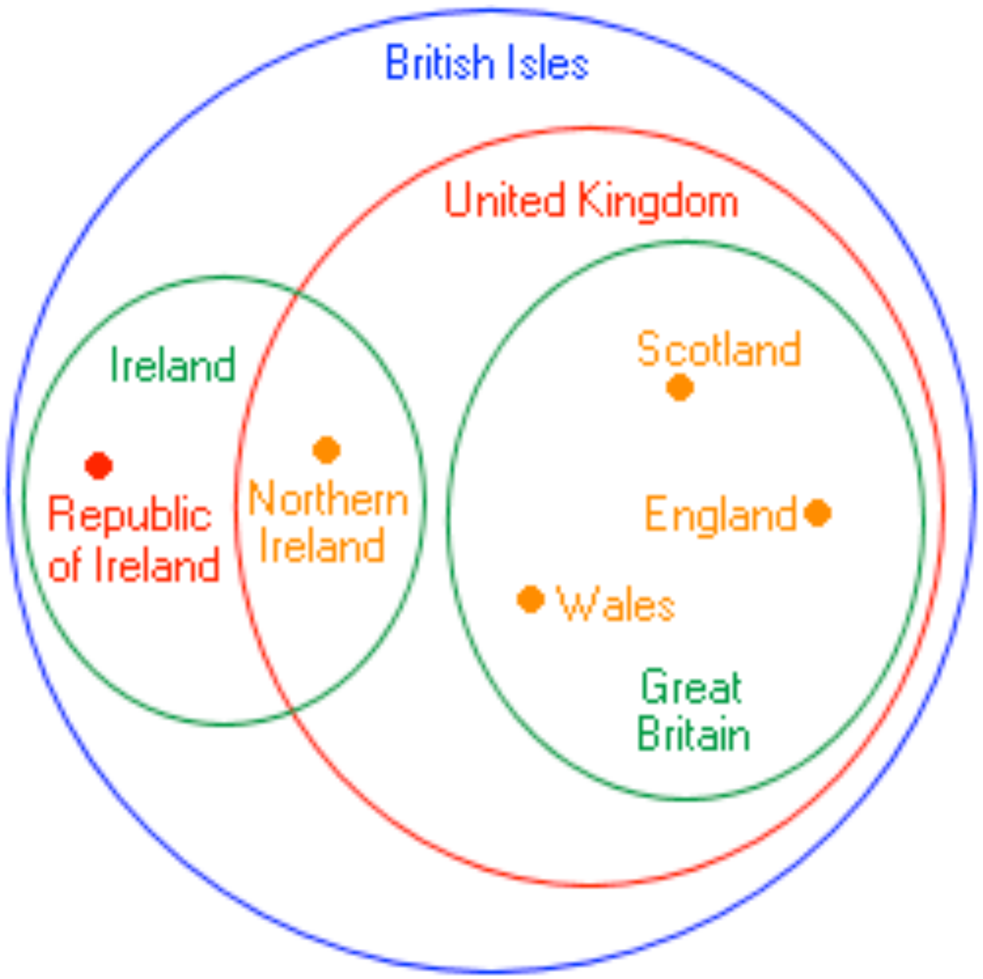


EULER

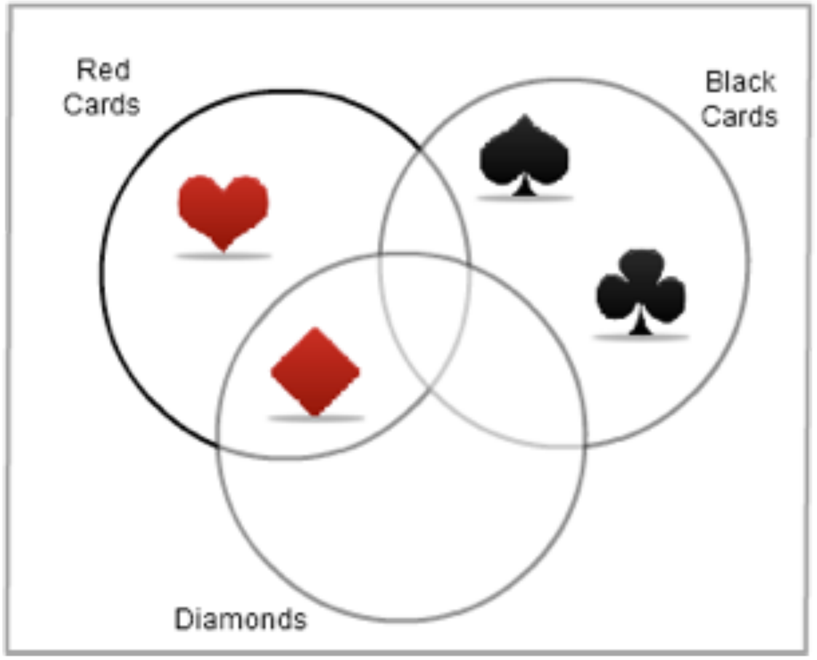


euler diagrams

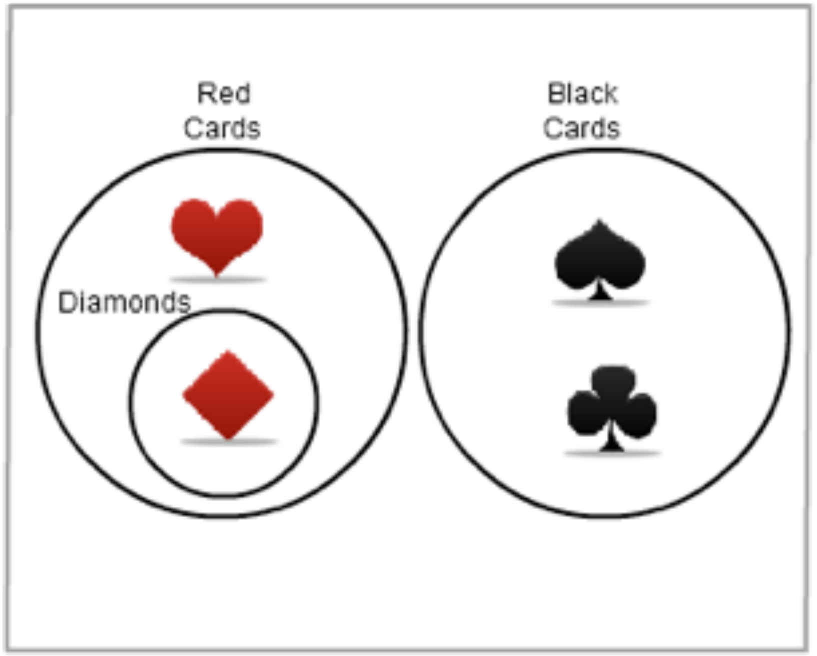
- show only existing relationships



VENN

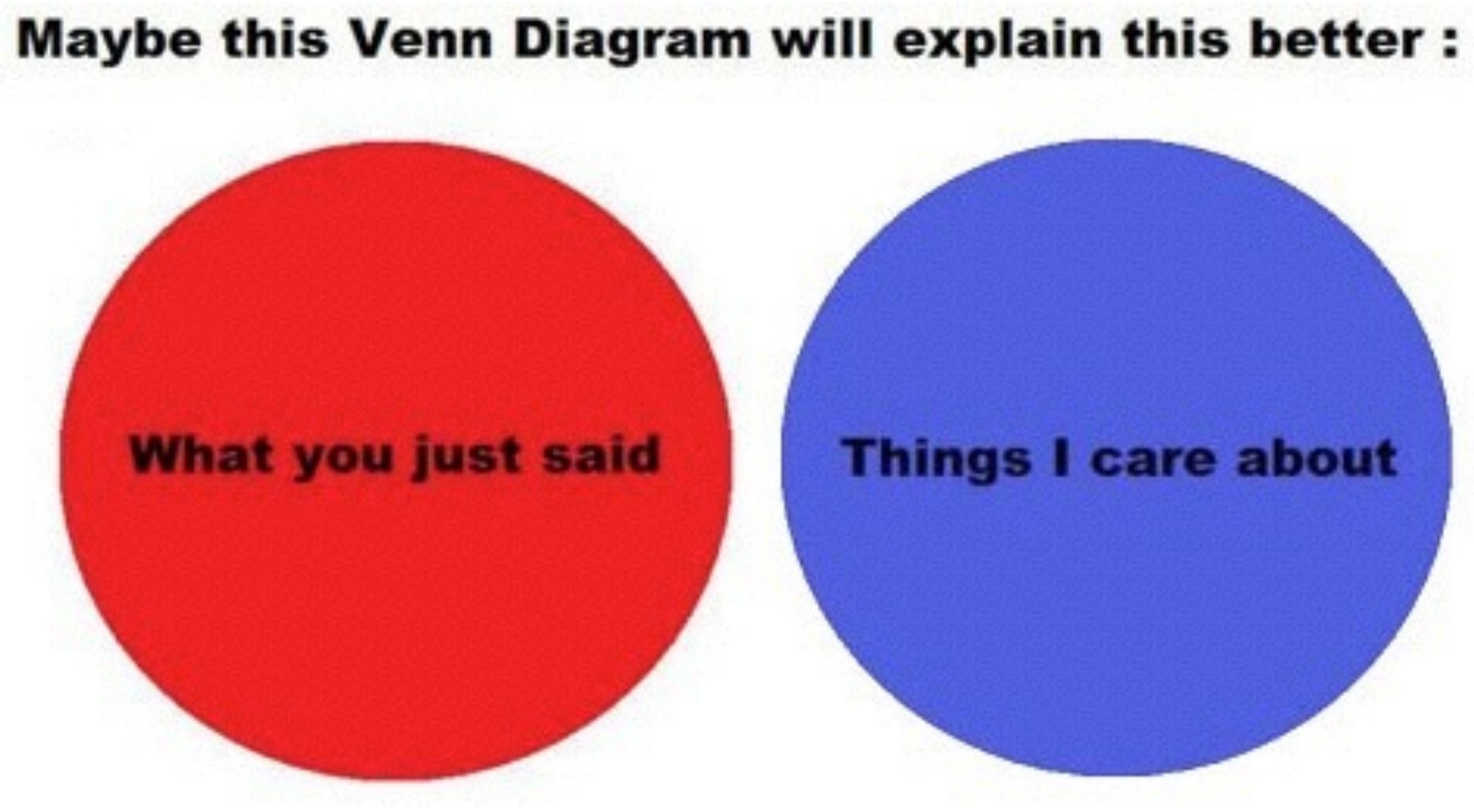


EULER



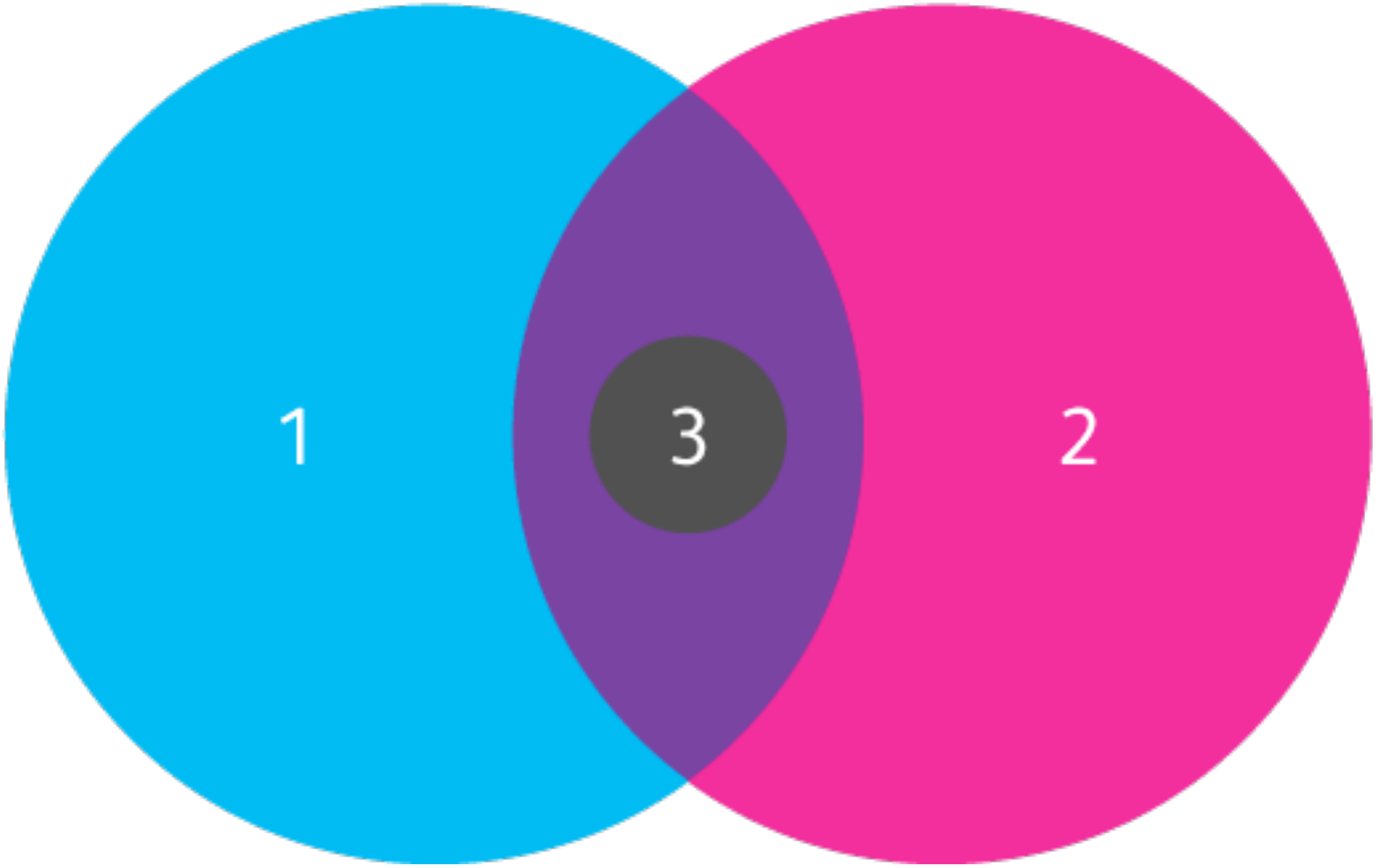
euler diagrams

- misunderstood



euler diagrams

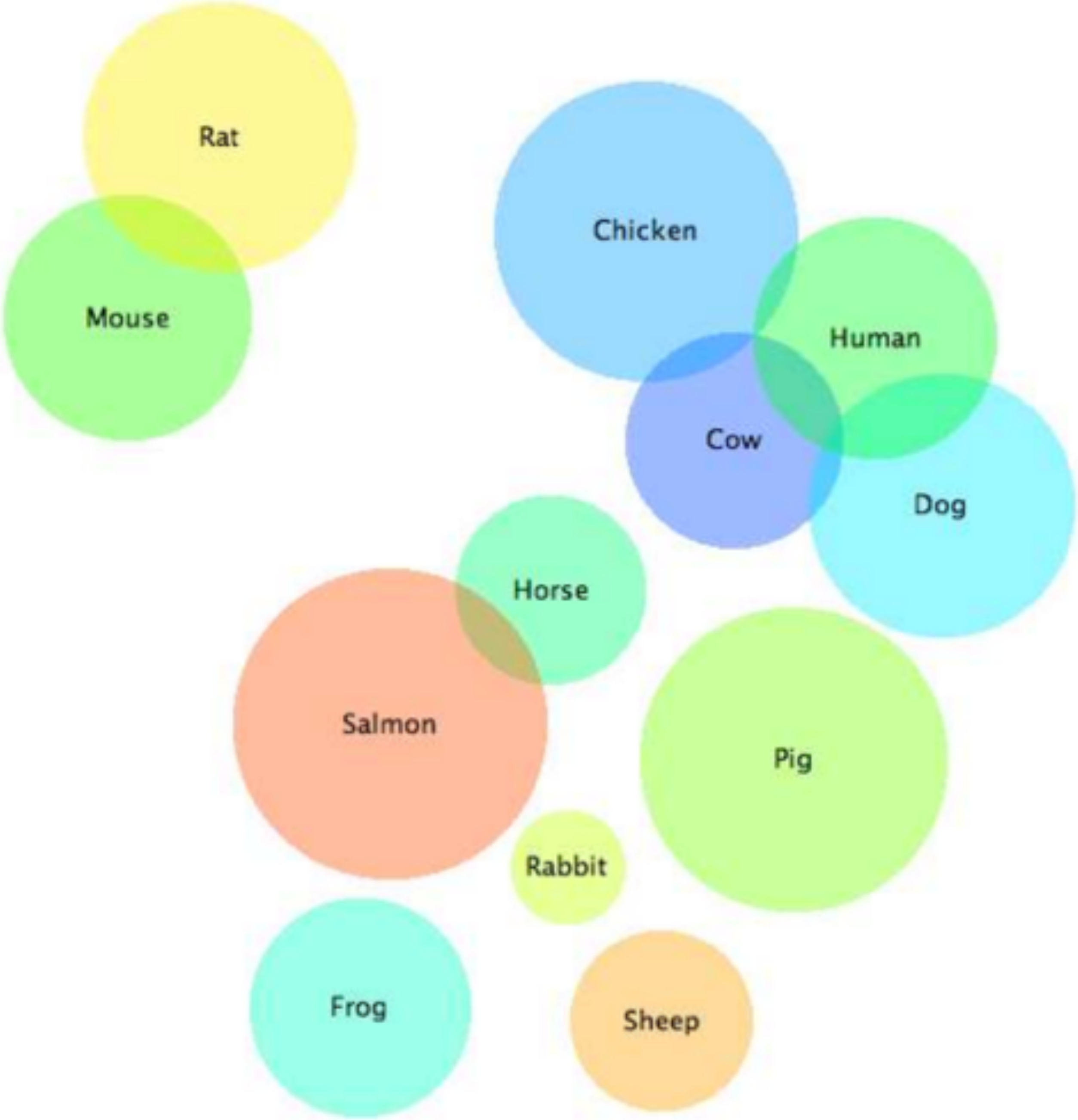
1: People who know what a Venn Diagram is.
2: People who know what an Euler Diagram is.
3: People who know the difference.



venn & euler diagrams

- adjust for area

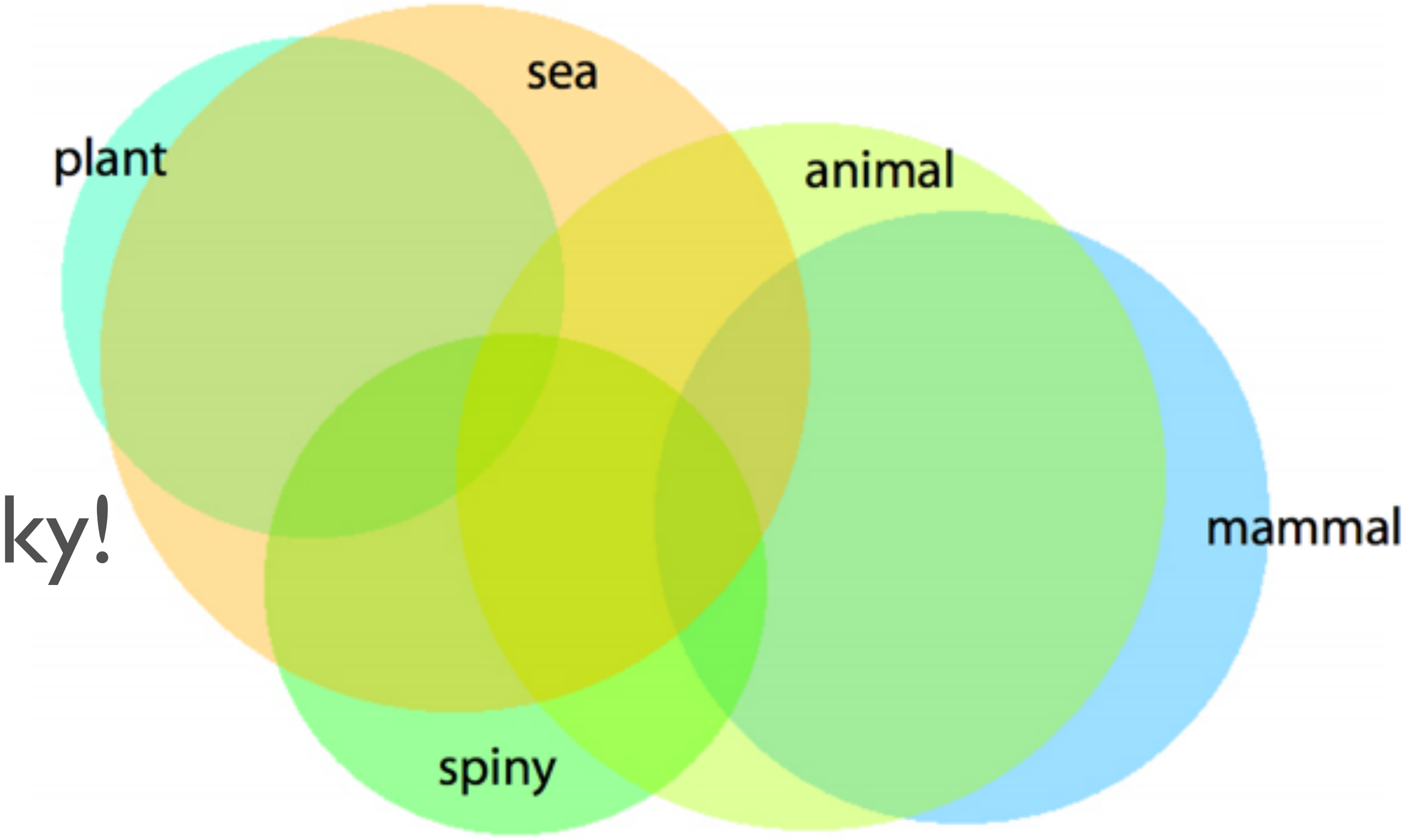
- starts getting tricky!



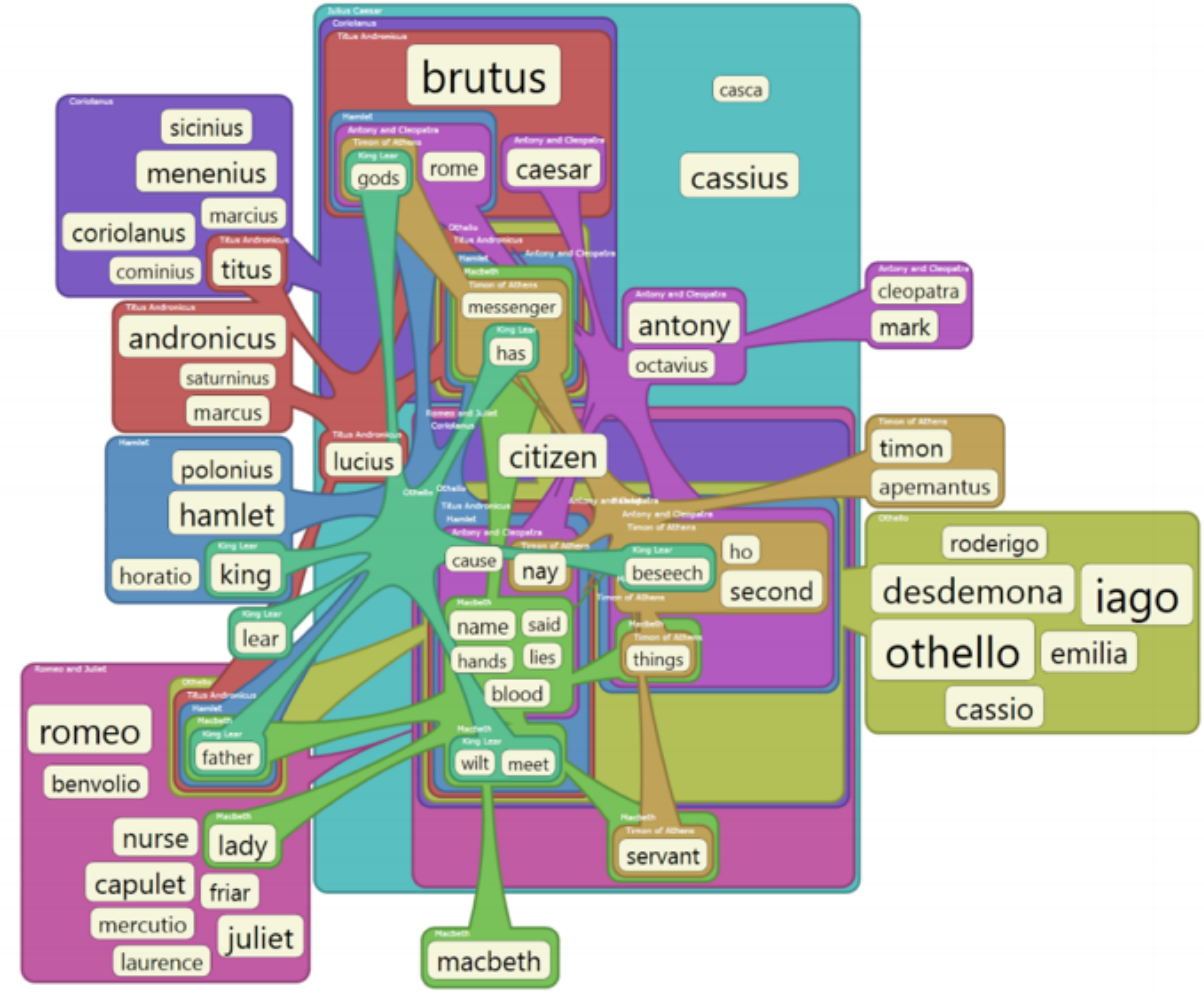
venn & euler diagrams

- adjust for area

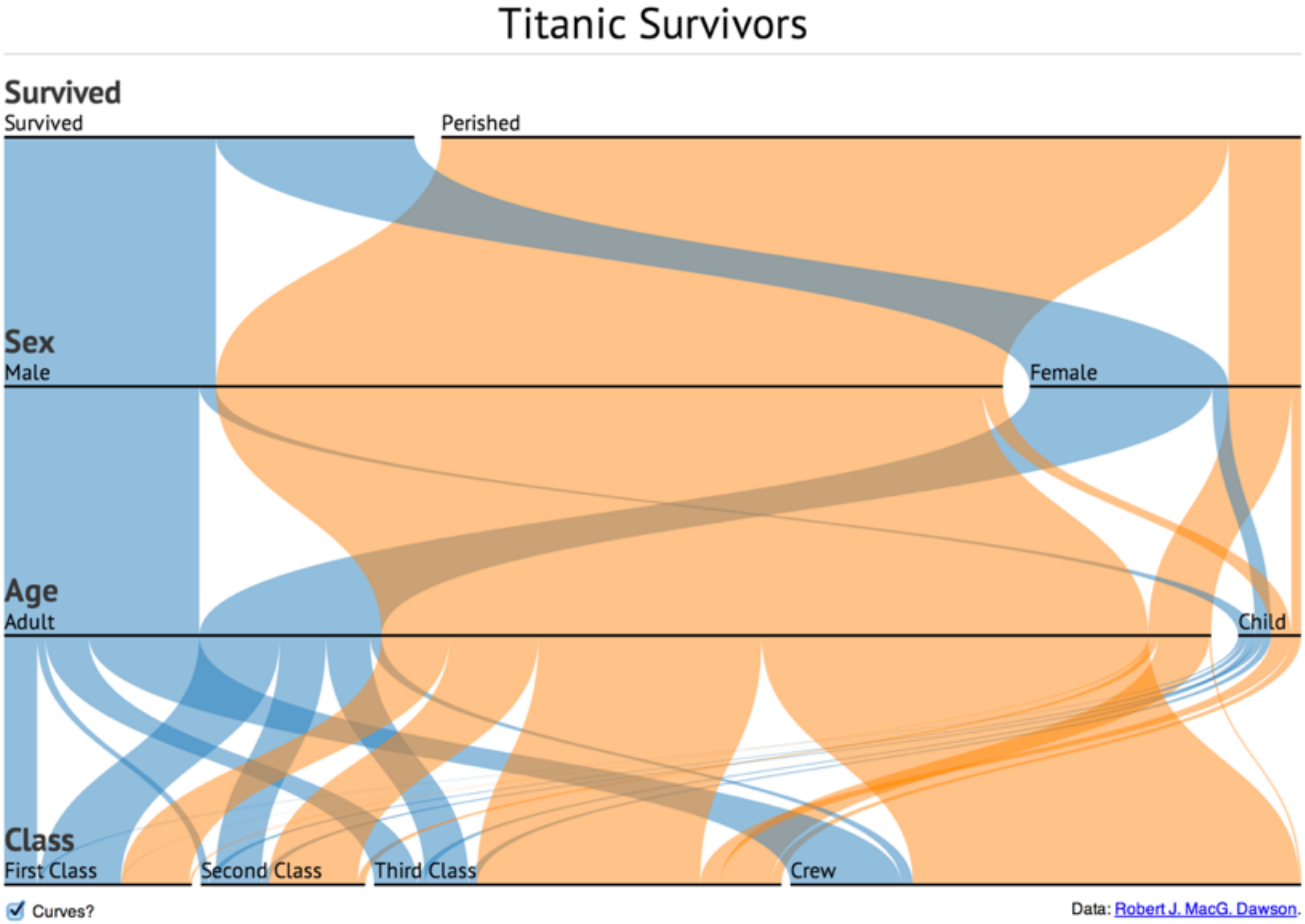
- starts getting tricky!



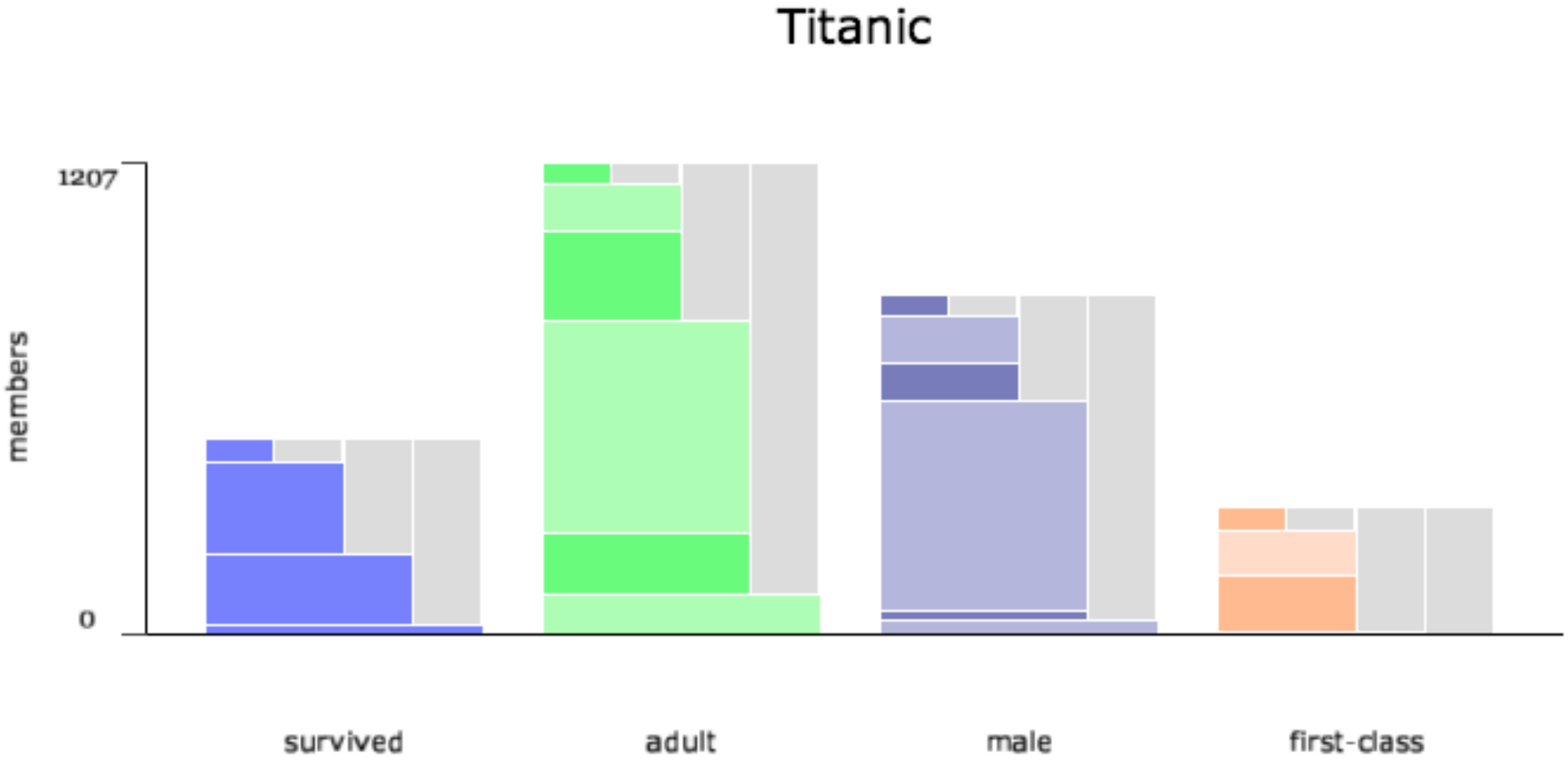
compact euler diagrams



parallel sets



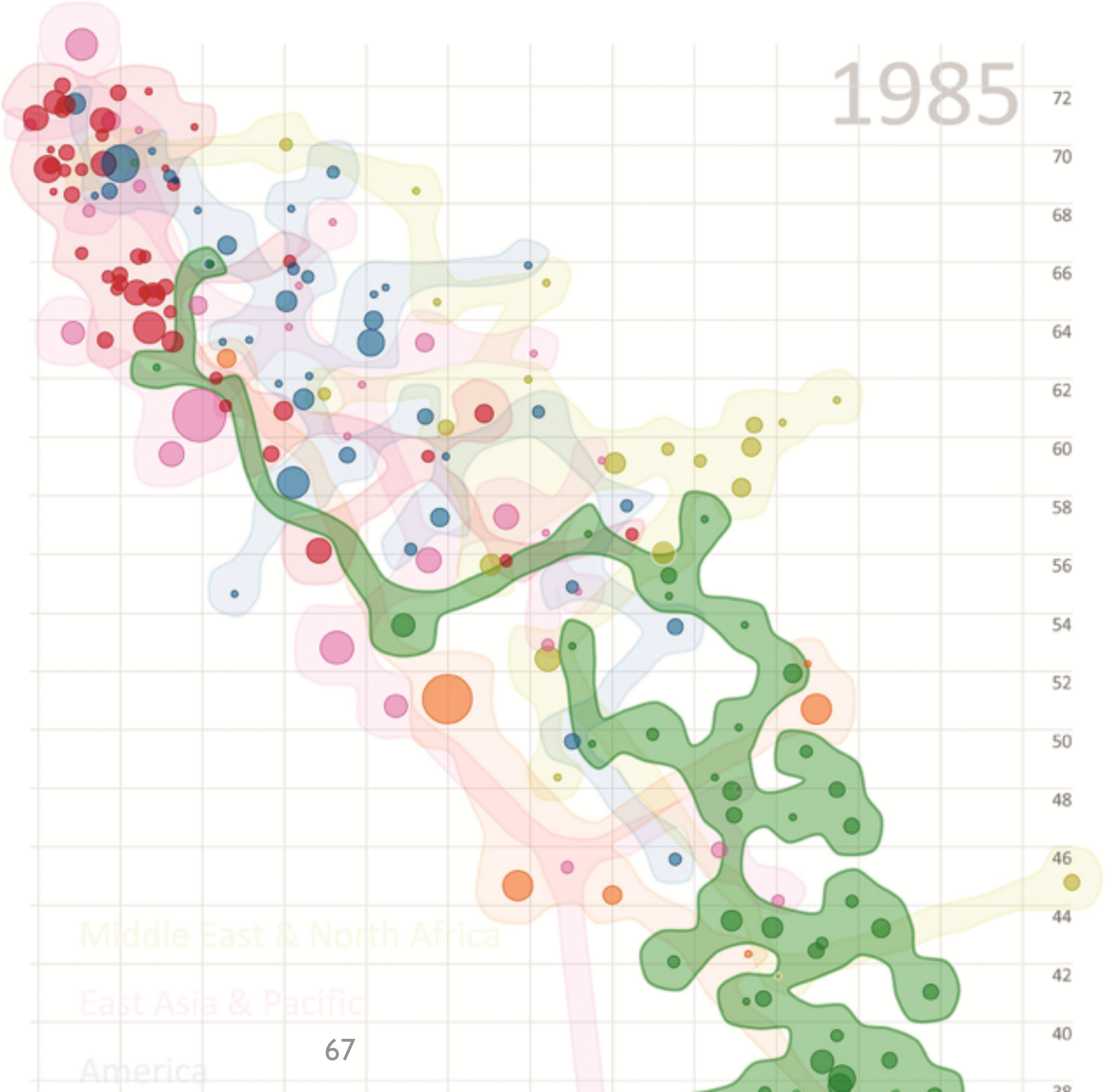
set o'gram



visualizing sets with constraints

bubble sets

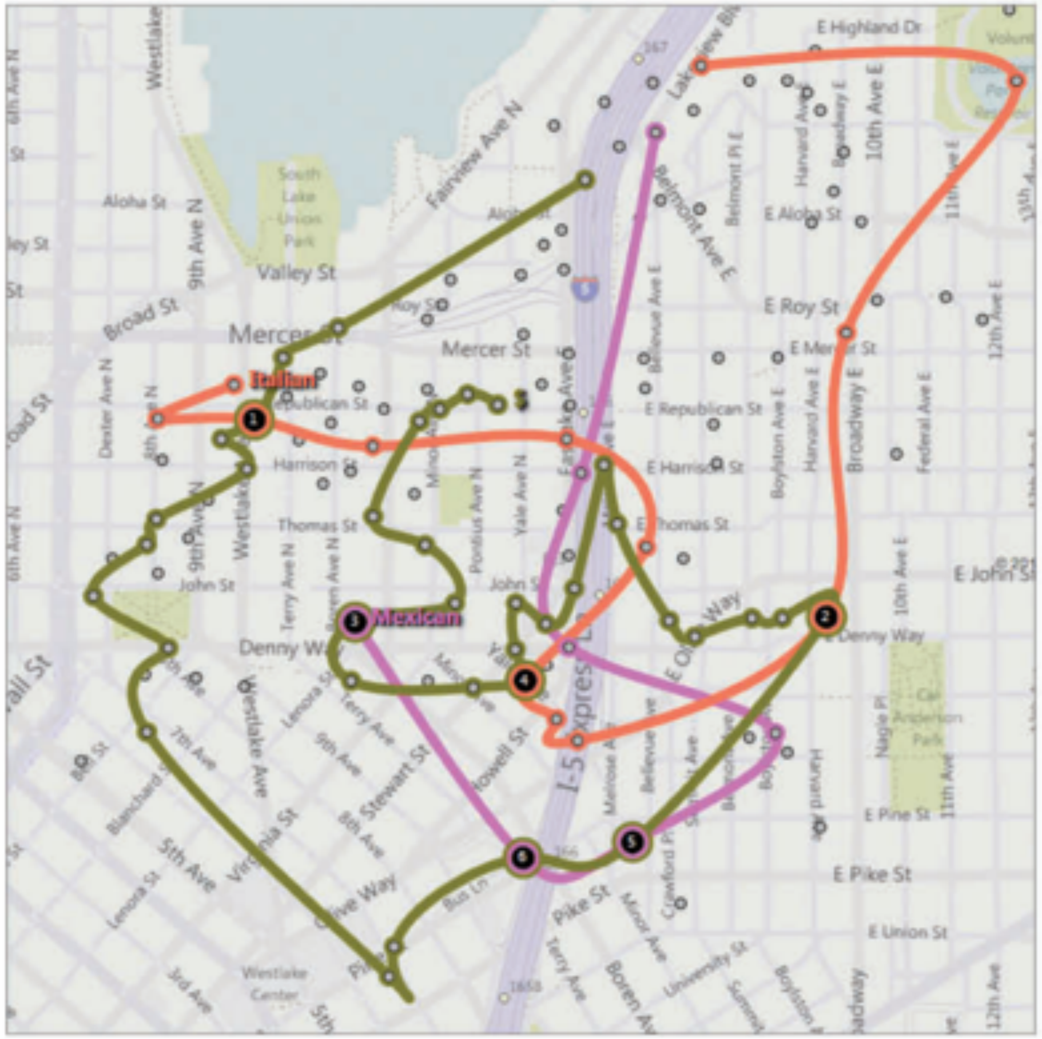
- connect points

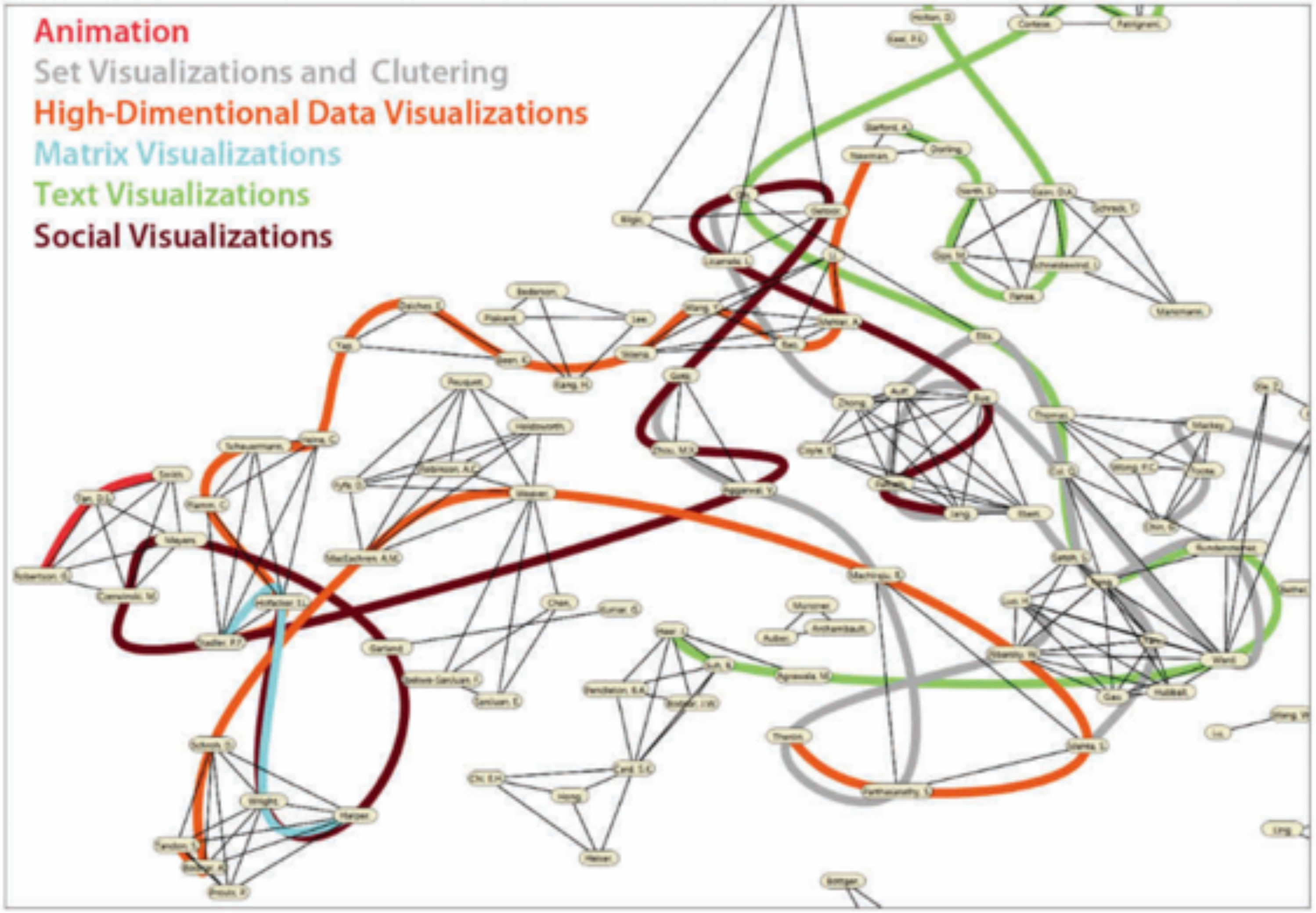


line sets

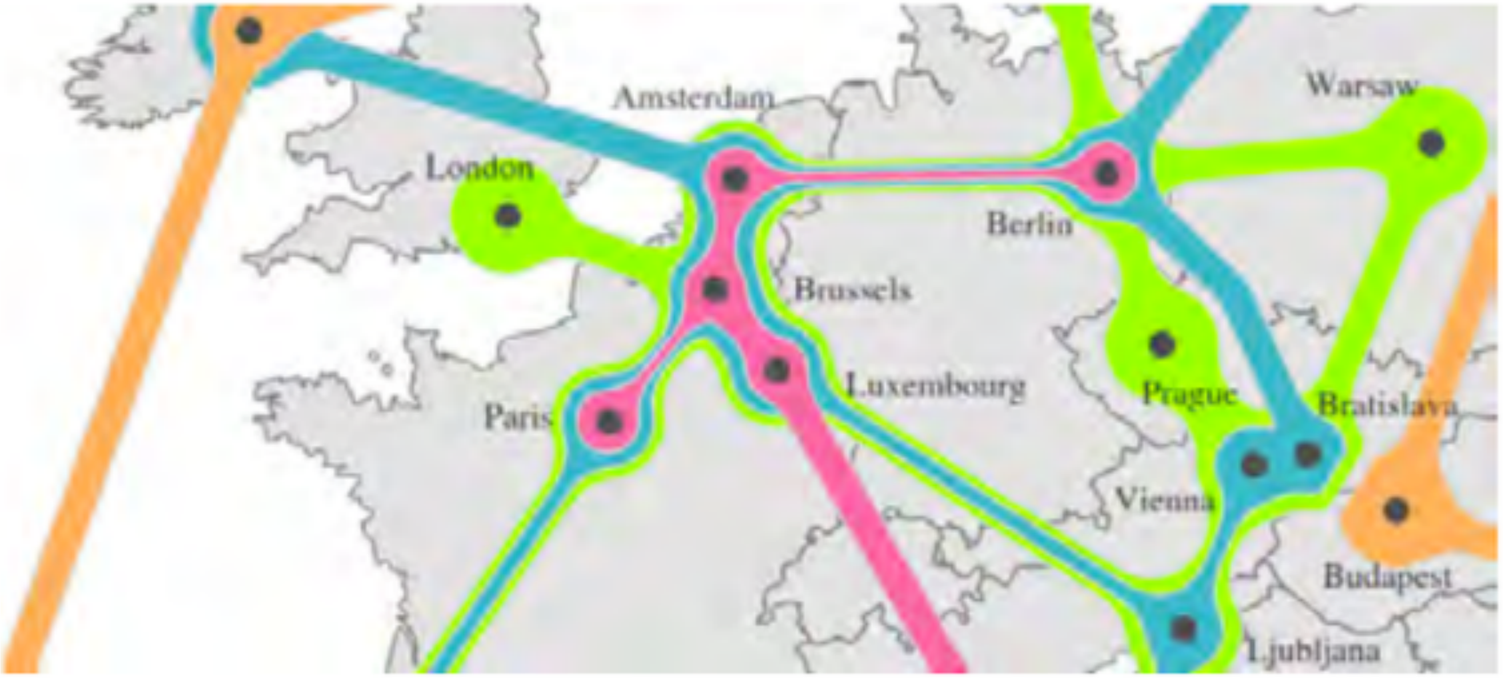
- restaurants

social communities



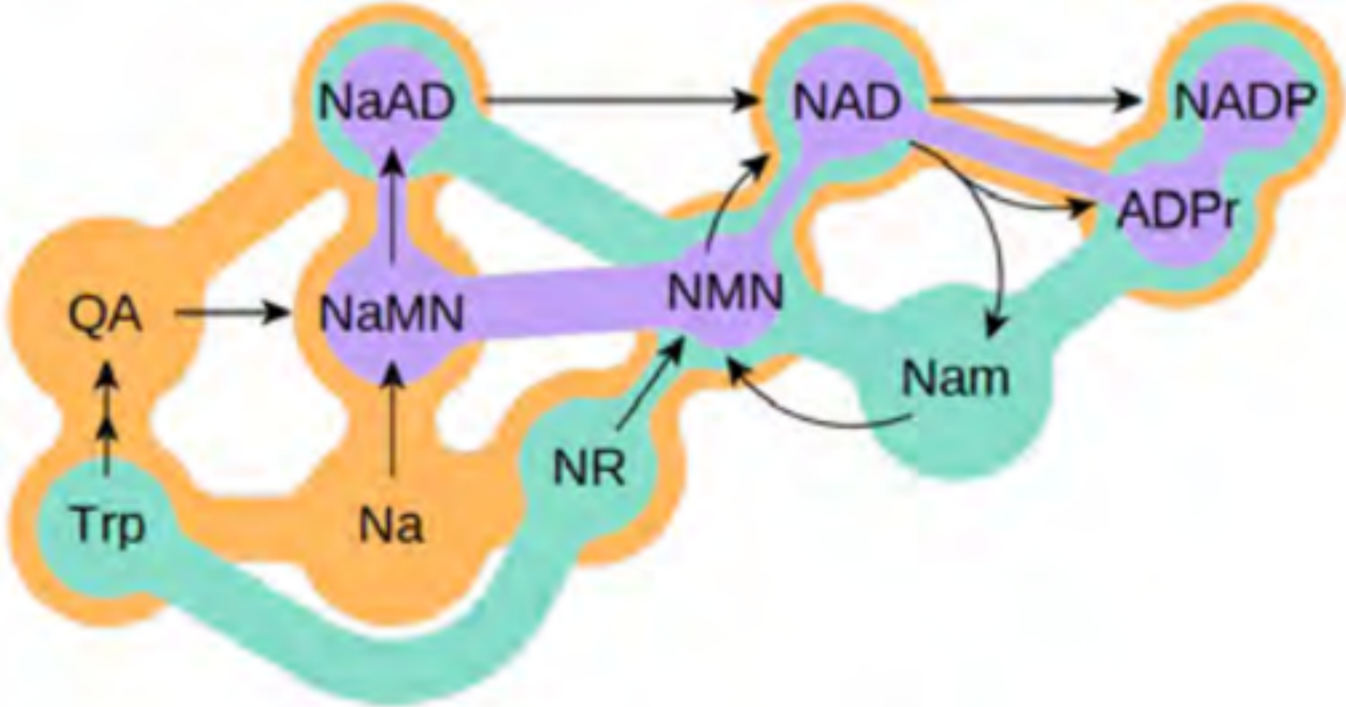


kelp diagrams



- cities on a map

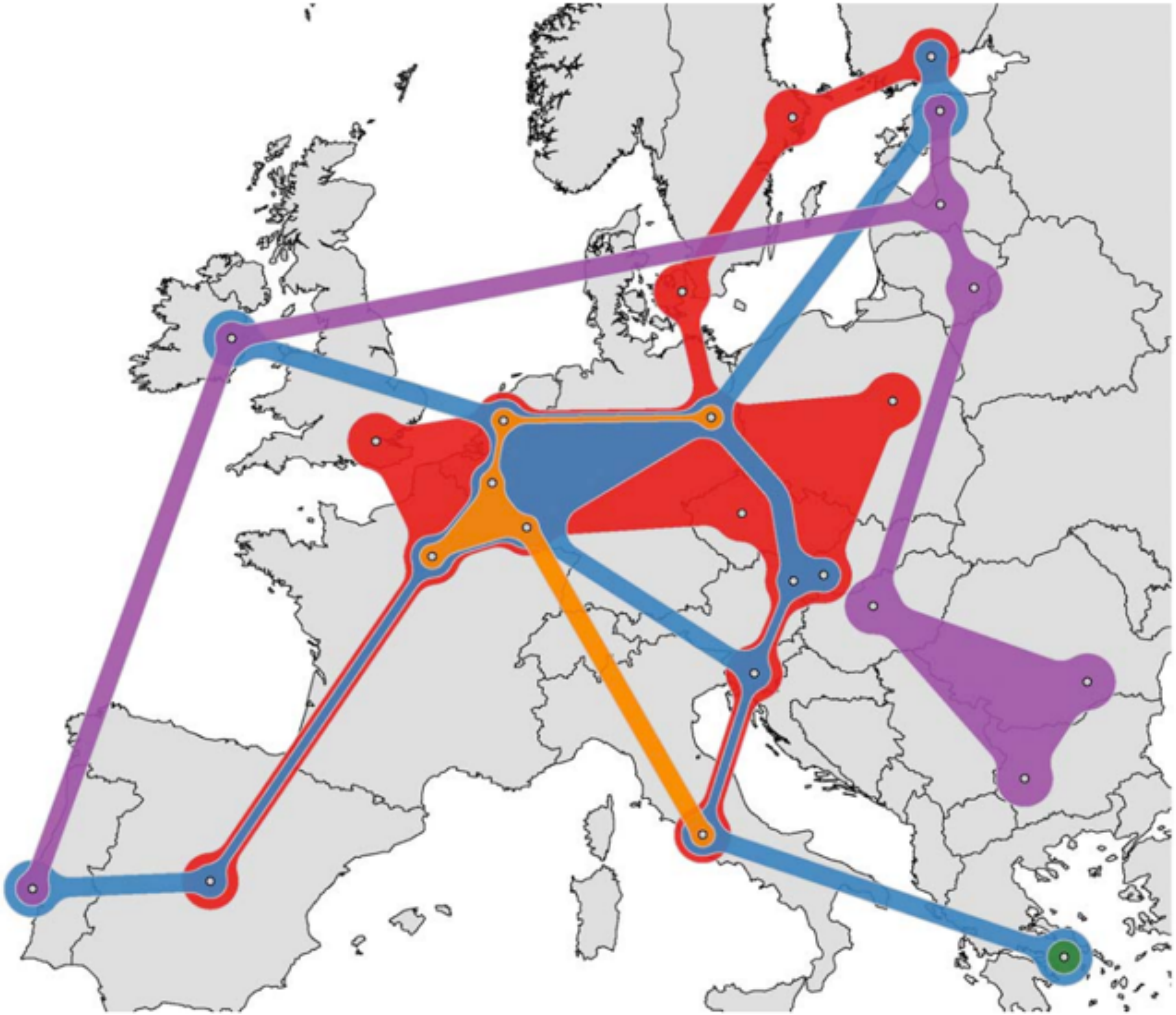
- metabolic network



kelp fusion

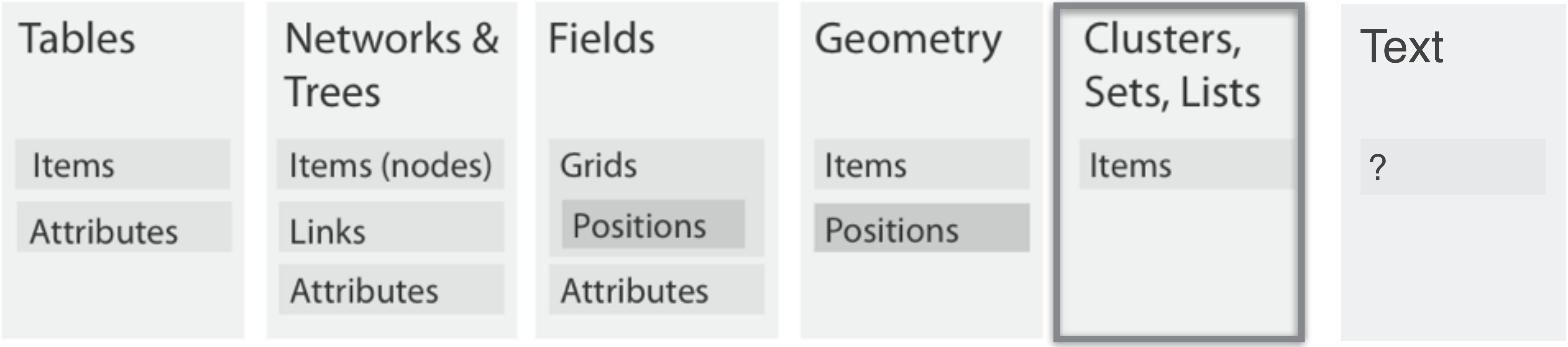
- cities on map

- lines & areas



sets

- applies to many datasets
- many combinations may be interesting
- limited numbers of sets



Tables
Items
Attributes

Networks & Trees
Items (nodes)
Links
Attributes

Fields
Grids
Positions
Attributes

Geometry
Items
Positions

Clusters, Sets, Lists
Items

Text
?

L15: Maps

REQUIRED READING

Chapter 8

Arrange Spatial Data

8.1 The Big Picture

For datasets with spatial semantics, the usual choice for *arrange* is to *use* the given spatial information to guide the layout. In this case, the choices of *express*, *separate*, *order*, and *align* do not apply because the position channel is not available for directly encoding attributes. The two main spatial data types are geometry, where shape information is directly conveyed by spatial elements that do not necessarily have associated attributes, and spatial fields, where attributes are associated with each cell in the field. (See Figure 8.1.) For scalar fields with one attribute at each field cell, the two main visual encoding idiom families are isocontours and direct volume rendering. For both vector and tensor fields, with multiple attributes at each cell, there are four families of encoding idioms: flow glyphs that show local information, geometric approaches that compute derived geometry from a sparse set of seed points, texture approaches that use a dense set of seeds, and feature approaches where data is derived with global computations using information from the entire spatial field.

8.2 Why Use Given?